A. Introduction and challenges for research workforce development:

A.1. Vision and goals/Intellectual Merit. The CI-TRACS program responds to the need for Climate Science (CLS) research workforce development in the area of CI. Without sophisticated CI skills CLS in the Hawaii-Pacific region is fundamentally limited with concomitant impacts upon our ability to monitor, predict and mitigate the impacts of climate change on island locales. The **vision** of CI-TRACS is a CI-empowered climate science research workforce pipeline that transforms Hawaii-Pacific CLS across academic research, agencies and community stakeholders. The **goal** of the program is to build a multi-level capacity enhancement program for CI training and skills development that serves CI users, contributors and professionals in the CLS research community.

CI-TRACS aligns with **Solicitation Goals** of *Broadening Adoption of Advanced CI; and Integration of CI Skills into Curriculum/Instructional Material Fabric.*

A.2. Intellectual Merit.

Climate Science research in Hawaii. 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. *Priority:* As climate change continues to increase temperatures in the tropics, Hawaii and other Pacific Islands are at the forefront of the negative impact of these changes. Ocean warming leads to thermal expansion and glacial melt which may result in sea level rise of 3 feet by the end of the century [Anderson 2015]. Ocean warming is increasing the number and severity of tropical cyclones impacting the Pacific, creating substantial risk for island communities [Chu 2020]. Warmer oceans lead to ocean acidification, more frequent and extensive coral bleaching events and destruction of coral communities and the biota they support [Hoegh-Guldberg 1999]. Changes in climate have decreased annual rainfall for over two decades impacting native species and terrestrial ecosystems, increasing the threat and frequency of wildfires with increasing consequences for local agriculture and food security [Trauernicht 2019]. Sea level rise threatens coastal communities though salt water inundation, fouling of fresh water aquifers and disrupting coastal ecosystems [Fletcher 2020]. Chronic flooding exacerbated by extreme weather events threatens sovereign nations in the Pacific with rich cultural heritages defined by their homelands. *Opportunity:* The Hawaiian Archipelago is the ideal natural laboratory for tropical terrestrial ecology research, with a multimillion-year chronosequence of substrate age, diverse soils, a large elevation (temperature) range, steep moisture and energy gradients, and approximately 15,000 known terrestrial species [Eldredge and Miller 1995] including an estimated 1,400 native vascular plant taxa.

Examples of high impact terrestrial ecology research that exploit the experimental potential of Hawaiʻi's climate gradients include: weathering rates and soil formation [Sowards *et al.* 2018], biogeochemical cycling in soils [Siegenthaler *et al.* 2020; Helfenstein *et al.* 2018; Selmants *et al.* 2016; Chadwick *et al.* 2003], aboveground ecosystem carbon density [Asner *et al.* 2016], gross primary production [Kimball *et al.* 2017], carbon sequestration [Selmants *et al.* 2017], temperature dependence of soil CO_2 efflux [Litton *et al.* 2011], plant physiology and stress tolerance [Barton *et al.* 2020; Cordell *et al.* 1998; Krushelnycky *et al.* 2020], climate effects on forest tree diseases [Hennon *et al.* 2020; Fortini *et al.* 2019; Morales *et al.* 2011], forest management and restoration [Rayome *et al.* 2019; Friday *et al.* 2015], the phytobiome [Amend *et al.* 2019], effects of climate on ecosystem structure [Asner *et al.* 2005], species composition, diversity, and turnover [Ibanez *et al.* 2018; Craven *et al.* 2019], climate-change-induced species range shifts and impacts on endemic endangered species [Germain and Lutz 2020; Fortini *et al.* 2017; Atkinson *et al.* 2014; Krushelnycky 2013; Rovzar *et al.* 2016], species invasion and its impacts [Bremer *et al.* 2019; D'Antonio *et al.* 2017; Ibanez *et al.* 2019; Asner *et al.* 2010], drought management [Frazier *et al.* 2019], wildfire risk and management [Trauernicht 2019], greenhouse gas emissions from wildland fires [Hawbaker *et al.* 2017], watershed and coastal conservation [Oleson *et al.* 2017], effects of land use on coral reefs [Carlson *et al.* 2019]. As part of global ecosystem analyses, Hawaiʻi plays an important role in studies of climate dependencies of species composition, diversity, survival, and turnover [Gei *et al.* 2018; Ibanez *et al.* 2018; Johnson *et al.* 2018; Menge *et al.* 2019] and analysis of changes in carbon exchange under warming [Gallego-Sala *et al.* 2018]. For these projects, spatial patterns and temporal variability of climate are essential environmental covariates driving biotic and abiotic processes and patterns. Obviously, these are all data-intensive computationally-intensive research areas that can be empowered through CI infrastructure and skills in the research workforce.

Training and skills gaps for climate science in Hawaii. 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 (CLS) at UH and in the Pacific Islands through the adoption of advanced CI skills, technologies and access to national resources. Currently there is 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. 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 overwhelming number of CI tools, science gateways and data resources available for climate research and choosing the appropriate tools is challenging and confusing. Many software packages evolve rapidly and 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.

Inclusion of under-represented groups. Pacific islands are sentinel locations for the effects of climate change. In a sense, the Pacific today represents a global tomorrow. The effects of climate change are being felt in loss of ecosystem services from bleached coral reefs, seawater inundation effects on freshwater supply and soil nitrification, leading to depopulation, health and social inequity, trauma and erosion of traditional cultures. Indigenous Pacific cultures have historically strong kinship relationships with land, ocean and weather. However, the severe under-representation of these groups in the contemporary STEM endeavor limits Pacific self-determination of climate futures, Native Hawaiian Pacific Islander (NHPI) access to environmentally-related job opportunities and critically weakens climate science through a lack of indigenous voices and perspectives. The disenfranchisement of NHPI from STEM encompasses both ethnic under-representation, and other categories of persons such as low income, rural and formerly incarcerated individuals (the latter is critical to consider given the over-incarceration of NHPI). In the COVID era, these populations also include unemployed and under-employed persons as a result of the economic distress faced by Hawaii and further amplified in the US -affiliated Pacific. As CLS becomes more computationally intensive the dearth of CLS/CI and DS (Data Science) focused educational opportunities from K-12 onwards in Hawaii and the Pacific will manifest as a further digital divide and limit participation of NHPI in environmental research and natural resource management. Simply put, while NHPI seek environmentally-focused careers which are culturally-sustaining and of tremendous value to their communities, they will be less competitive if CLS/CI training gaps are not mitigated.

A.3. Prior experience and expertise in CI education and training.

This CI-TRACS *Implementation: Medium* program builds on strong foundations in CI capacity builds (infrastructure, human capital) and piloting of educational and training programs in the period 2013-2020 at UH and its partner institutions. This pilot phase of education and training program development has been sufficiently extensive, evaluative and reflective that it justifies the current proposal's positioning at the implementation level of NSF's CyberTraining program.

Capacity Builds: HI-DSI, the UH LAVA lab and the UH CI program.

In 2013 Gwen Jacobs joined UH as the Director of Cyberinfrastructure and led the growth of CI with institutional investments in HPC and a series of NSF grants in workforce, software engineering, international networking, and multidisciplinary research in the EPSCoR Ike Wai Project (#1557349) Program. UH was among the inaugural group of institutions in the ACI-REF project (#1341935) which developed best practices in research computing facilitation. Two consecutive awards in international networking through the NSF IRNC program (#1451058, #2029312) support research and education networks worldwide. Sean Cleveland led UH in the AGAVE (#1341935) and now Tapis (#1931575) software framework development in collaboration with TACC. In 2014 Jason Leigh began construction of LAVA, the Laboratory for Advanced Visualization & Applications, to provide world class visualization cyber-infrastructure for the UH System. This laboratory was primarily funded by awards from NSF, including a CI award for SAGE2 (#1339772), and an MRI award for building CyberCANOEs (cyber-enabled collaboration analysis navigation &

observation environments- award #1530873) in the UH System (in total 18 were built in 3 years with a combination of funding from NSF and other sources). In 2018 Jacobs and Leigh launched the Hawaii Data Science Institute (HI-DSI) [HI-DSI] with startup funding from the President of UH and the Vice Chancellor for Research. Since its inception, the HI-DSI has facilitated the hiring of 2 new data science faculty in Information & Computer Sciences (ICS) and Hawaii Institute for Marine Biology. With funding from EPSCoR, four new data science faculty were hired at UH Hilo. HI-DSI has received 4 cyber-infrastructure awards (Tapis #1931575) for collecting and processing streaming data, (SAGE3- #2004014, 2003800) for supporting collaboration in the cyber-enabled scientific research and development enterprise, PIREN (#2029312) 100G networking from UH and islands in the Pacific to US mainland and Australia, and MANA (#1920302) the UH computing cluster now with both CPU and GPU nodes and high performance storage. In 2020, University of Hawaii at Hilo completed construction of their visualization laboratory to enable them to more closely collaborate with UHM's research efforts. In 2019 CUH completed its Data Science Center (physical facility) and launched major, minor and certificate programs in Data Science with 4 new faculty including two NHPI. CyberCANOES (SAGE-enabled visualization infrastructure pioneered by Leigh) are now at UHM, CUH, UHH, UHWO, Kamehameha Schools, Hawaii State Energy Office (Pacific locales) and UH West Oahu's Academy for Creative Media. There are currently ~4000 SAGE2 users located at ~800 sites in over 17 countries worldwide. Each year for the past 10 years we have held a Birds of a Feather meeting and live demonstration at Supercomputing, to introduce users to SAGE's newest features. For SAGE3, which incorporates AI into SAGE2, we will expand these presentations to conferences that are more specific to the study of climate science (such as the American Geophysical Union, Earthcube) to introduce these audiences to what SAGE3 can offer in understanding the global climate change problem.
CI Expertise: Leadership and the CI Team. PI Jacobs has been active in CI research and national policy for over 30 years, in neuroinformatics, data management and campus and national cyberinfrastructure efforts. She served as member and Chair of the NSF Advisory Committee for CI (2016 - 2020) and co-author of the NSF CI2030 Future Cyberinfrastructure report [CI2030]. Co-PI Leigh has been conducting research, development and deployment of CI for more than 20 years in NSF funded projects such as GeoWall (visualization hardware for 3D viewing of geoscience data- award #0219246), CoreWall (software for annotation & visualization of lacustrine, ice and ocean core drilling data- award #0601978), SAGE, SAGE2, SAGE3- software for cyber-enabled collaboration and visualization, and NetSage (#1540933)- project to collect and monitor NSF research network traffic. The CI Team built by Jacobs and Leigh deployed to the CI-TRACS program comprises software engineers, HPC and science gateway specialists. Joint appointments/roles across the HI-DSI-EPSCoR-UHH-CUH continuum (Stokes, Turner, Jacobs, Cleveland, Leigh, Giambelluca, Tipton, others) facilitate collaboration.

Complementary Funded programs and partnerships. The CI-TRACS program sits at the interface of numerous funded programs and partnerships in addition to the CI grants above. UH and CUH have partnerships with the Texas Advanced Computing Center (TACC); AGAVE (OAC 1450413) TAPIS (OAC 1931575) and the INCLUDES CUH summer DS program HRD 1744526). PI Jacobs and Co-I Cleveland lead UH partnerships with the Science Gateways Community Institute [SGCI], Campus Champions [CHAMP] and Campus Advanced Research Computing Consortium [CARCC] The Hawaii EPSCoR program (OIA 1557349, PI Jacobs, co-PIs Leigh and Turner) currently supports major science gateway, HPC and CI infrastructure for its hydrology research focus. CI-TRACS co-PI Turner leads a suite of diversity and inclusion STEM programs: NSF AISL (DRL 1811691), I-USE (DUE 1525884), 2 S-STEMs (DUE 1833772 and DUE 2030654) and the CUH HHMI 'Inclusive Excellence ' program. Turner is co-PI on the Pacific wide 'Islands of Opportunity' NSF LSAMP (HRD 1828684) which has recently initiated DataCamp certifications and research projects under the leadership of CUH and UHH.

Pilot work in development of CI training experiences. For the last 6 years, our team has experimented with a variety of different learning approaches to develop training programs focused on three sets of skills - basic CI training, basic data science skills and advanced data visualization and collaboration. These training programs span formal coursework, hands on CI workshops targeting faculty, grads and state agency professionals and immersive summer bridge programs for undergraduates. Since 2014 we have served 1,000 researchers in more than 110 HPC training and 30+ CI and DS training events. The CI-TRACS leverages what we have learned from these programs and is designed to meet the needs of three populations of participants: those with little to no experience, graduate students engaged in climate research with limited experience with specific tools and advanced training for faculty and research professionals in state agencies. We summarize outcomes and identify lessons learned for these programs below that inform the proposed CyberTraining Program.

(a) Formal coursework: UH's ICS (Information and Computer Sciences) department has formal courses in Machine Learning, Data Science, Artificial Intelligence, and Data Visualization, which are all heavily attended. **Lessons Learned:** 1. Although many domain science students are interested in taking the class, there is considerable trepidation in taking 4 semester-long classes while also attempting to satisfy their own graduation requirements. 2. As classes are all project-based, there are opportunities for the domain science students to propose their areas of research and data as candidates for the projects; 3. The data is more readily available and the domain science students serve as subject matter experts for the ICS students; 4. The ICS students are able to create a prototype that the domain science students can continue to use after the class; 5. The resulting projects are also valuable to ICS students as they are able to showcase practical applications of their knowledge on their resumes; 6. As the landscape of data analysis tools and techniques grows it has become increasingly difficult for faculty and students to stay abreast of ongoing innovations. A valuable sub-project was to have every student choose a new tool to learn and present to the class. That knowledge is then codified into a database to create a growing and evolving survey of tools.

(b) Hands on workshops in CI skills. Since 2014 we have trained more than 1,000 researchers in more than 110 HPC training and 30+ training events. These events provide users with basic and advanced training in CI skills - including HPC, software development and scripting, scientific workflows, data management, data visualization, virtual reality, and advanced collaboration tools. ITS-CI uses the best leading practices from ACI-REF [ACIREF], Campus Champions [CHAMP], and CARCC [CARCC], and leverages partnerships with Jetstream 2 [JETSTREAM2] and the Texas Advanced Computing Center (TACC). Past workshops included introductory HPC concepts and usage, advanced computation topics such as GPU accelerators, domain specific topics such as Genomic workflows and data management with Agave [AGAVE], reproducible science with containers and Jupyter notebooks with TACC. Within the HI-DSI, we have three certified Software and Data Carpentry instructors and provide all of our training materials on GitHub. We also have an HPC quick-start guide [ITSCI1] and YouTube channel [ITSCI2] for self-paced training and review. As part of the systemwide activities of the HI-DSI we will expand these training activities across institutions via CyberCANOE and Zoom to increase participation from all campuses. The ITS CI team has recent experiences organizing and running virtualized workshops through service as conference chair (Jacobs), technical chair (Cleveland) and visualization chair (Leigh) during the Practice and Experience in Advance Research Computing 2020 July conference (PEARC20), an event with over 1,000 participants and 27 tutorial/workshops and 60 presentations, has prepared this team for coordinating large tutorial/workshop and presentation sessions on-line.

(c) Undergraduate summer bridge programs.

(i) Hilo Data Science Summer Bridge (NSF EPSCoR #1557349). Four new tenure track faculty in data science hired through this award designed a five week summer bridge program that supported 51 high school and undergraduate students at UH Hilo from 2017 to 2019. Students began with little to no experience in data science and participated in an intensive bridge experience which gave them a credit-bearing Mathematics course, an Introduction to Data Science course and Fundamentals of Data Science in R course. Students learned a range of techniques to include using neural networks to automate detections in songbird recordings of an endangered species endemic to Hawaiʻi Island obtaining and utilizing machine learning techniques to optimize educational games, data collection focused on computational and statistical analyses of coral health [Burns 2020]. **Metrics:** 60 students over 3 years, 57% female, 45% URM, 8% NH. Varied majors (Marine Science, Computer Science and Mathematics). **Lessons learned:** A flipped classroom approach, paired with utilizing the online assessment portal MyMathLab on the CyberCANOE allowed both the Precalculus and Calculus 1 courses to be taught in parallel. In Summer 2018, the bridge program was updated to include a credit-bearing Fundamentals of Data Science in R course and Machine Learning.

(ii) NSF INCLUDES (#1744526) Chaminade University Data Science Summer Immersion. The SPICE (Supporting Pacific Indigenous Computational Excellence) summer immersion is a 1 month summer program. **Metrics:** 45 students in 2019 and 2020 (80% female, 96% URM, 89%NHPI). This one month immersion assembled participants into small learning communities who self-selected into thematic areas (health inequity, Covid 19, climate change, social justice) and performed specific research projects within those thematic areas (e.g. COVID-19 data, Pacific fisheries data, seawater inundation data, teen pregnancy data, food deserts), culminating in a capstone video presentation [link]. **Lessons Learned:** We identified three themes based on SPICE 1 (2019) experiences and designed programming to address them in SPICE 2 (2020). These will inform the design of CITRUS. ***Increase coding self-efficacy.*** SPICE 1 students, especially non-CS or -DS majors, lacked confidence in their abilities stemming in part from lack of coding and basic vocabulary in DS. *Response:* Two week pre-SPICE experiences (online DataCamp modules curated by faculty) to build basic competencies in all participants and demystify the DS field. In SPICE 1 we identified weaknesses in project ideation and ability to design a testable hypothesis. SPICE 1 was essentially an open model where we asked students to fully conceive their own project on any topic. *Response:* In SPICE 2, students self-sorted into hui around specific research topics. During the DataCamp bootcamp and week 1, peer mentors, faculty and students worked aggressively to define a workable project and related datasets, which was finalized by week 2. These learning communities struck an improved balance between independence and inexperience with a more prescribed project that still allowed for student ownership. ***Promote sense-of-belonging***. Indigenous students do not necessarily have a sense of enfranchisement into computational disciplines, which can limit participation and engagement. Responses: In SPICE2 we invited 4 SPICE 1 alumni, who were Native Hawaiian (NH) women, as peer mentors and role models. Together with NH faculty and our STEM Cultural engagement Specialist, they implemented programming focused on role modeling and their mentorship was equally privileged with that of non-Hawaiian faculty. Inclusive pedagogy and content presented indigenous epistemologies/values/knowledge application as assets to DS. ***Achieve project ownership.*** Some students in SPICE 1 demonstrated behaviors that we associated with passivity and lack of a sense of project ownership. *Response:* In SPICE 2 we selected thematic project areas (hui themes) that linked tangibly to community, culture, regional priorities and place. We also provided, through the video showcase, opportunities to demonstrate learning and project outcomes to family and community members, which is a culturally-consistent approach.

(d) Undergraduate skills development program. CUH-HI-DSI (Stokes) designed, developed, deployed and evaluated the 'Curated DataCamp' skills-development model. This provides for democratized participation (DS/CS majors, as a value-add to non-CS majors, certifications for community participants for upskilling and reskilling) with high completion rates of skills training modules in DS. **Metrics:** To date four groups have completed the curated DC model at CUH: (1) 30 Native Hawaiian STEM Scholars who are non DS majors (certificate of completion for resume building and professional development); (2) 45 SPICE participants (some DS majors, Biology, Environmental, Business and Criminology majors certificate of completion for resume); (3) 11 Participants from NSF LSAMP network (Turner, co-PI HRD 1826864) from Palau CC, American Samoa CC, U. Guam, College of the Northern Marianas, College of the Marshall Islands, Hawaii Pacific University UHH, College of Micronesia), 2 semester program with Fall 2019 KSA training and Spring 2021 research project on locally-relevant data sets); (4) 2 working professional participants who are epidemiologists for American Samoa Dept .of Health (professional development, DataCamp model completion followed by COVID-19 and Rheumatic Heart disease data analysis projects). *Overall completion:* 1292 student-module hours with a 92% module completion rate. **Lessons Learned:** High completion rates derived from elements of the curated DC model: faculty work with students to identify an initial prescriptive pathway of modules according to prior knowledge, interest and goals; faculty provide technical assistance throughout modules; faculty supervise a mini-capstone research project based on a culturally-, career- or community-based data set; data products are curated in Github and in form of a video presentation that can be shared with employers, family and community stakeholders.

(e) Sustained engagement of undergraduates. An overarching lesson learned from multiple undergraduate research enrichment programs across UHM, UHH and CUH is the benefit of intentionally supporting longitudinal engagement of students after the summer immersion is ended. In CI-TRACS, the CyberClub is intended as a sustained community of practice and vehicle for longitudinal engagement (much like IEEE and ACM student chapters). But whereas the IEEE/ACM chapters are primarily for undergraduates, our CyberClub will consist of both graduate and undergraduate members. In summer immersions and CyberClub we will be intentional about building community through social and cultural activities alongside academic interaction, and we will emphasize peer mentoring and peer support, as well as graduate to undergraduate mentoring. CI-TRACS will strive to act as a conduit connecting summer immersion undergraduates to (a) mentored paid research opportunities with climate science faculty, (b) assist students in finding mentors for their required senior capstone classes within the CI-TRACS network, (c) publicize opportunities for mainland REU and similar internship programs that are relevant to the CI/CS-focused CI-TRACS undergraduates. Furthermore, as a student-run organization they are eligible for financial support from the university.

B. Description of the training program

The previous section outlined the concurrent but disparate efforts in bringing about Computational Science, Data Science, and Cyber-infrastructure training to our stakeholders. CI-TRACS proposes to leverage the lessons learned from those past efforts to create a new highly streamlined, scalable and sustainable CI-training model for the UH System aimed at addressing climate science research problems, but potentially generalizable to other scientific disciplines. The modern scientific research enterprise and data lifecycle comprises project conceptualization/brainstorming, data collection & cleaning, data analysis & visualization, knowledge crystallization, and knowledge dissemination & presentation, all of which can be enhanced through CI [Leigh2019]. The CI-TRACS program will provide 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.

B.1. Activity 1. Broadening Adoption of CI skills.

B.1.1. Activity 1A. Academic Year CI Fellows Program.

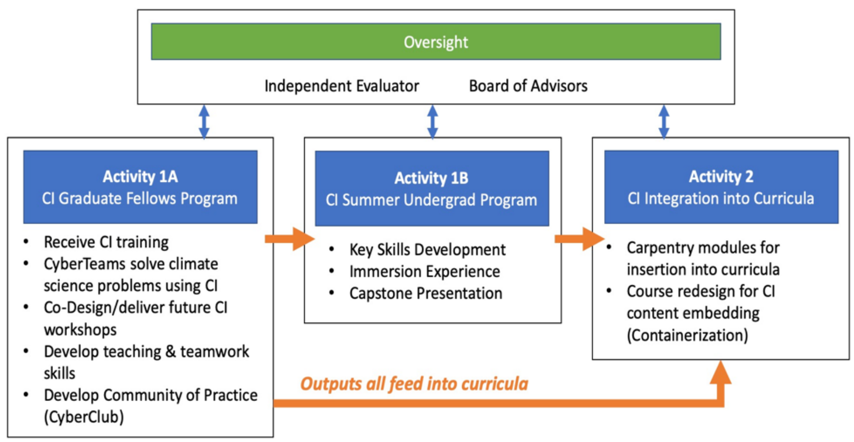

Figure 1. CI-TRACS Overview

(a) Recruit and train CI Fellows. The academic year program is designed to meet two needs: 1) focused immersive workshops in CI skills for climate science researchers and 2) broad training for CI Fellows who will participate as trainers in the workshops. CI Fellows will be recruited from departments (CLS, CS and DS) with the goal of achieving a balance between CS and DS students and those in atmospheric, geo and ocean sciences. In addition to their training activities they will work collectively in small CyberTeams of 2 on climate change problems related to thesis research. Professional development activities will be programmed jointly with current HI-DSI Data Science Fellows Program [HI-DSIFellows].(e.g., portfolio development in GitHub; speakers/discussion on STEM inclusion and bias, speaker/discussion on research ethics, data sovereignty, mentoring/career development panels, job application workshops, grant writing basics). *Proposed annual timeline:* *Summer:* Call for applications/nominations (May 1), Application Deadline (June 1), Selection and Notification (June 15); *Fall Semester:* Research projects, workshops and CyberClub*; Spring Semester:* Research projects, workshops and CyberClub, *Summer Semester:* Mentors for CITRUS students, Research projects, CyberClub.

(b) Design and deliver CI workshops. During the academic year, senior personnel and CI Fellows will design and deliver a set of 10 hands-on workshops focused on a set of topics designed to span the research lifecycle of climate science projects, starting with problem conceptualization to data dissemination and archiving. These workshops will target the sets of basic and advanced skills necessary for climate scientists working across three broad areas: atmospheric science, hydrology and ecology. The workshops are described below in Section B: CI Literacy

(c) Develop teaching and teamwork skills. Each cohort of Fellows will be selected for a range of skills and expected to work as a team and leverage their complimentary skills. They will play a major role in the development of the workshop curriculum. As described below in the workshop design section the fellows will migrate existing workshop content to the workshop format designed by The Carpentries community [Carpentries] and create lessons focused on specific skills with climate science use cases and datasets. The cohort of CI Fellows will gain skills in CI, pedagogy and collaboration.

(d) Develop a Community of Practice - the CyberClub. In addition to their formal training in CI and as trainers in the workshops each cohort of Fellows will participate in a "CyberClub" modeled after the UC Berkeley Blueprint Tech for Social Good program [BLUEPRINT]. Fellows will serve as ambassadors for the CI-TRACS program, maintain the collection of training materials and community data sets and serve as consultants for individual researchers with CI challenges in their research. As a team, they will apply their CLS research and newly acquired CI skills to work with stakeholders in state agencies and with non-profit organizations to solve CI challenges in data management, analysis, visualization and dissemination. These activities will hone skills in multidisciplinary collaboration and identifying CI solutions to common challenges.

B.1.2. Activity 1B. CI Summer Undergraduate Program.

The CUH-HI-DSI-UHH team (summer program lead: Stokes) will implement a CI training program based on the INCLUDES-funded SPICE DS summer immersion program (a 4 week bootcamp) and incorporating lessons learned both from SPICE and the UHH Summer DS Bridge program. The CITRUS program (**C**yber**I**nfrastructure **TR**aining for **U**ndergraduates in **S**ummer) is designed to be held online synchronously (as we did for the INCLUDES SPICE program successfully in 2019/20) but additional in-person elements could be added if conditions permit. Program elements are as follows: *(1) KSA (Key Skills and Abilities) Training*. Using the faculty-curated online modules model developed in SPICE, participants will be prescribed a suite of essential training modules from various (e.g., Software Carpentries Core, DataCamp modules, CUH-ITS CI training modules) and mentored through their completion by faculty and CI graduate fellows. (a) A pre-kickoff subset base of modules (e.g. essential coding modules, Creative Thinking; Data Wrangling; FAIR Data Management, Security and Ethics; Data Visualization) will be assigned in the two weeks prior to the start of the 1 month CITRUS immersion and additional modules will continue to provide skill building throughout the CI immersion. These additional modules will be prescribed as needed by program faculty. *(2) Research Immersion*. Four week online immersion using a synchronous online format. Elements include weekly orientation, mentored co-working sessions, co-work within Climate Science Hui teams (Hui is the Hawaiian word for group), independent work, guest lectures on key topics, social events, ethics and data sovereignty training, training and activities around bias reduction and inclusion. Weekly schedule: *Week 1*; Identifying projects, questions, group members, moving, cleaning, organizing data. *Week 2*; Complete cleaning and organizing datasets, exploring and analyzing datasets. *Week 3*; Analyzing and creating visualizations. *Week 4*; Finishing project and work on final presentations. Daily schedule: *8:30.* Morning Check in and community building activity. *9:00*; Roundtable project discussion/updates/challenges to learning community. *9:30*; Project work in themed groups with peer mentors (CI Fellows, proper CITRUS alumni in years 2 and 3). *12:00*; Lunch break/Lunch seminar. *13:00*; Training modules/Project work. *16:00*; Project reflection with learning community or Seminar/Training/Discussion panel slots. *16:30*; Homework assignments with faculty/peer mentors (training modules or project work). Close. *17:00*; Faculty and mentor debrief or Learning Community Social Activity. *Capstone presentations.* Video capstone presentations will be recorded in week 4 and presented to invited stakeholders on the Friday of week 4.

B.1.3. Exemplar CITRUS and CI Fellows research projects.

Fellows and CITRUS students will engage in climate science research projects leveraging the expertise and data of our faculty advisors. Projects will be multifaceted, collaborative and will focus on reinforcing CI skills in a real world climate research project. Co-PI Giambelluca will lead this effort working with climate science colleagues across UH. Table 1 shows examples of possible projects that can be adapted to the immersive 4 week CITRUS and 1 year CI Fellows programs. We will aim to have a network of at least 10 collaborating faculty.

Table 1. Example researchers and project overviews.
Giambelluca, Tom (UHM). *Next-generation atmospheric, hydrologic, and ecological research enabled by statewide Cyber-Infrastructure for real time climate data.*
Synopsis: National climate data compilations and analyses include inadequate coverage of the Hawaiian Islands. As the only US tropical state Hawaiʻi is a model system for functional responses of ecosystems to changes in atmospheric CO2. **Student Goals:** Create automated workflows for ready access to a spatially comprehensive, high quality, reliable climate data set and data analysis products covering the Hawaiian Islands. Design and implement a hypothesis-driven project using these workflows (e.g. response to rainfall and antecedent moisture conditions in diverse basaltic landscapes; interactive effects of temperature, moisture, solar radiation on ecosystem carbon fluxes). **Outputs:** Improved models in the atmospheric sciences, hydrology and ecology for Hawaii. **CI Content:** 1. Collect streaming data from the current Halenet and planned Mesonet monitoring system; 2. Automate data screening, homogeneity testing, and gap filling; 3. Automate analysis; 4. Integrate data science methods; 5. Data products: daily near-real-time, high-resolution, gridded digital map products and reports of climate variables (e.g., precipitation, temperature, humidity, wind speed, solar radiation) using containerized workflows scheduled on HPC or cloud resources; and data management and sharing services for dissemination.
Burns, John, (UHH). *Assessing coral heath with digital imagery and analysis.*
Synopsis: Climate change induced ocean acidification and temperature increases accelerate the frequency of large-scale mortality associated with disease outbreaks and mass coral bleaching events. There is a need for efficient methods using high resolution image analysis for detecting and monitoring coral health. **Student Goals:** Use digital analysis, machine learning to classify coral images and assess climate impact on different coral species. **Output:** Improved methods and data for classifying health impacts on coral communities **CI Content:** photogrammetry, data cleaning, image classification, machine learning, data resource of classified coral image data.
Sadowski, Peter, Torri, Guiseppe (UHM) *Predicting cloud cover to improve solar power reliability.*
Synopsis: Using machine learning to identify cloud patterns and now casting predictions to determine the amount of sunlight reaching solar panels that generate electricity for Hawaii. **Student Goals:** Train an algorithm to 'learn cloud patterns' using existing satellite data, and predict when and where clouds are expected to appear. **Outputs:** Systematic prediction of cloud cover with enough advance to adjust power grids. **CI Content:** image data analysis, ML, atmospheric modeling.
Crow, Susan, (UHM). *Climate change impacts on soil carbon storage.*
Synopsis: Understanding soil carbon can help the agricultural sector reach sustainable growth, climate change resilience, and climate change mitigation. What metrics comprise the critical factors necessary to develop an organic matter-based soil health index for Hawaii's diverse, tropical soils? **Student Goals:** Using the Soil Incubation Database, analyze distribution of soil carbon content across Hawaiian Islands and compare to mainland agricultural areas. Create visualizations of results. **Outputs:** Comparative analysis between regions will bridge the gap between site-level measurements, and global remote-sensed data or data products derived from models aimed at assessing global-scale rates of decomposition and C turnover. **CI Content:** accessing data, time series analysis, R statistics, GIS mapping, visualization.
Tsang, Yin Phan (UHM). *Impacts of climate change induced streamflow changes on native fauna.*
Synopsis: Climate change impacts the tropical water cycle, a critical driver of freshwater ecosystems. A significant decline in dry season flows indicates that more streams may become intermittent, which has important implications hydrological connectivity, and management of Hawaiʻi's native migratory freshwater fauna. **Student Goals:** Identify data sets, analyze for changes in distribution of biota, create GIS maps correlating streamflow with loss of native species. **Outputs:** GIS visualizations of changes in distribution and abundance of species with changes in stream flow: **CI Content:** data management, cleaning, R analysis, GIS visualizations.

B.2. Activity 2. CI literacy: Integration of CI skills into curriculum.

The CI-TRACS program is designed to supplement the academic curriculum for undergraduate and graduate students through hand-on immersive workshops and to contribute a set of curriculum materials for inclusion in formal coursework to a central CI-TRACS GitHUB repository [CITRACSGithub]. The workshops will generate a rich set of training materials that can be incorporated into undergraduate or graduate courses. They will follow the Carpentries model [Carpentries] to insure consistency, clarity and enhance sharing of materials. Students will recognize and appreciate the familiar format of the lessons, and faculty can exchange information with each other more easily. CI faculty and specialists in the CI-TRACS program will also generate curriculum resources that provide training in CI basic skills and promote appreciation of the importance of CI to Climate Science research. These modules are envisaged as deployable to undergraduate and graduate classes in climate-related fields. Their embedded nature will alleviate expertise gaps in faculty who are not CI experts who teach these classes and promote CI awareness, skills and appreciation in both faculty and students.

B.2.1. Activity 2A. Workshop development.

(a) Workshop design: Workshops are designed following The Carpentries curriculum model to provide consistency in the pedagogy, organization of materials and delivery of content. As in the past, we will leverage existing peer reviewed Carpentry lessons and learn from and contribute to the Community Developed Lessons [CommunityLessons] in the area of climate science. We will contribute all new lessons (including data sets) generated through this project to the Carpentries Incubator, which will facilitate sharing as well as collaboration with colleagues at other institutions.

Table 2. CI Workshops, Target CI Skills and Learning outcomes	**Content source**
Problem conceptualization / Brainstorming	
W1: Creative thinking	
Synopsis: Introduction to SAGE 3 and how it can be used across the entire scientific research enterprise. **Learning Outcomes: (1)** Demonstrate ability to install, configure, and use SAGE3 in scenarios with example data and information relevant to climate change science; (2) Demonstrate participation in CI-enhanced brainstorming, data collection, data analysis. **CI Tools:** SAGE 3.	CI-TRACS will develop
W2: Scientific Workflows & Gateways	
Synopsis: Introduction to science gateways for climate science including the Hawaii Climate Data Portal, Ike Wai Gateway, GeoEDF, etc for access data and running computational workflows. **Learning Outcomes.** (1) Demonstrate understanding the advantages and disadvantages of various workflow models, and determine which is most appropriate to their current research workflow. (2) Demonstrate ability to take exemplar climate science data and apply it to one or more appropriate gateway systems. **CI Tools**: Hawaii Climate Data Portal, Ike Wai Gateway, GeoEDF and Hydroshare.	CI-TRACS will develop
Data collection, wrangling and management	
W3: Scientific Software Basics	
Synopsis: Foundations and essentials of scientific software. **Learning Outcomes:** Demonstrate ability to use the Unix command line, GitHub software version control and Anaconda software package management on the Mana HPC resources. **CI Tools:** Git and Anaconda.	Software Carpentry Core Lesson [SoftwareCarpentry]
W4: Data Wrangling with Computational Notebooks	
Synopsis. Introduction to the Python and Pandas for cleaning and wrangling data climate datasets using Jupyter notebooks on the Mana HPC for exploratory research and reproducible analysis. **Learning Outcome**: Demonstrate the ability to take exemplar climate change data and build Jupyter notebooks on Mana to conduct cleaning and wrangling of the data. **CI Tools:** Panda, R, Jupyter, Python.	Data Carpentry Core Lesson adapted from Ecology Curriculum and HIDSI resources [EcologyCarpentry]
W5: FAIR Data Management, Security and Ethics	

Synopsis: Introduction to FAIR data management practices. **Learning Outcomes:** Demonstrate ability to use the Hawaii Climate Data Portal, Ike Wai Gateway and Hydroshare for data annotation and dissemination while applying FAIR principles and practices in basic data security and ethics. **CI Tools:** GeoEDF, Hydroshare,GitHub, FigShare, Zenodo,	Adapted from extant HIDSI resources, Library Carpentry FAIR Data and Software [FAIRDATA]
W6: Data Movement, dissemination and archiving	
Synopsis: Introduction to different data transfer tools like ftp, rsync and Globus for moving data. **Learning Outcomes:** Demonstrate understanding of advantages and disadvantages of various bulk data transfer tools. Apply understanding to decide, based on their current computing, storage and network infrastructure and that of their end-point sites, appropriate solution for workflow. Demonstrate the ability to identify, communicate and mitigate potential bottlenecks in collaboration with campus cyberinfrastructure and network operators. **CI Tools:** SFTP, GridFTP, LFTP, RSYNC, Globus.	CI-TRACS will develop
W7: Smart Data Collection for sensor networks	
Synopsis: Introduction to the Tapis Streaming data APIs. **Learning Outcomes:** Demonstrate the ability to collect, store and manage time-series streaming data and employ data-event driven computation. **CI Tools:** TAPIS	Adapted from extant workshops PEARC20, Science Gateways20
Analysis, simulation & visualization	
W8: High Performance Computing	
Synopsis: Introduction to High Performance Computing on the Mana cluster via Open OnDemand to use interactive computing and the Slurm batch scheduler for running scientific codes. Introduction to NSF Jetstream2 cloud computing resource for using and creating virtual machines and the Tapis Abaco function-as-a-service for on-demand cloud computing resources to perform computation. **Learning Outcomes:** Demonstrate the ability to choose, access and employ HPC resources in support of scientific computation. **CI Tools:** Open on Demand, Jetstream2. Tapis	Adapted from extant HIDSI - HPC onboarding workshops
W9: Machine Learning approaches in Climate Science	
Synopsis: Machine Learning Techniques for spatial and temporal climate data. **Learning Outcomes:** Select, apply, and evaluate ML models. Describe data, models, and modeling assumptions. **CI Tools:** Open OnDemand, GPU computing, Scikit-Learn, Tensorflow.	Adapted from UHM ML and Data Science courses
W10: Data Visualization	
Synopsis: Introduction to Data Visualization for Climate Science datasets. **Learning Outcomes:** Apply exemplar climate science data to widely used Information (such as Plot.ly, Vega) and Scientific Visualization (such as Paraview, Visualization Toolkit) tools. Use SAGE3 to create mashups of their data, text and visualizations. **CI Tools:** CyberGIS, Plot.ly/Vega, ParaView, VTK.js, SAGE3.	Adapted from previous SAGE2 and SAGE3 workshops

The CI Fellows are integral to development, implementation and delivery of the workshop curriculum. In Year 1 (Fall), CI fellows (representing climate science and include computer science, data science, ecology, hydrology and atmospheric science) will work with the CI team to migrate previously developed workshops to a common format following the Carpentries model, and develop new modules/workshop content as necessary. Each workshop lesson will address a question or topic in climate science using data from local and national sources. Fellows will be tasked with identifying datasets and encouraged to choose topics aligned with their research interests. They will gain skills in basic programming and content management, GitHub, scripting, computational notebooks, dataset management, pedagogy and course design.

Training environments will be developed in containers (Docker or Singularity) for use with Jetstream2 resources. The training environments that include software tools and dependencies will be converted into images/templates and containers on Jetstream2 allowing re-use of the training environments but also allowing researchers to incorporate the training environments directly into their own research or their courses with no/low friction. MyBinder for Jupyter notebook environments will also be enabled for corresponding Github repositories allowing simple access to these resources for self-paced training. Table 2 shows the set of 10 workshops, their learning outcomes, CI skills taught and content resources. We will leverage existing content and workshop materials from our HI-DSI and Chaminade resources and will leverage the Carpentries Software, Data and Library Lessons in our workshop design. Climate science use cases and local data resources will be used, leveraging input from our faculty advisors.

B.2.2. Activity 2B. Develop and support CI curriculum/workshop modules that can be embedded into climate science courses by non-CI expert faculty.

(a) Course content. This activity is designed to embed CI content and training activities in a diverse range of courses without the need for extensive training of faculty (e.g. climate scientists) who are not CI experts. A suite of 'plug and play' resource sets will be developed and supported by CI-TRACS CI specialists and Fellows, for faculty to deploy in their courses in a low barrier fashion. These resource sets will use containers such as Docker and Singularity that allow for the encapsulation of CI software and library dependencies along with datasets and teaching materials applied to CLS disciplines. An invitation to climate science faculty will be generated annually with a goal of 3 course redesign collaborations per year. Course redesign and annotated resource set generation will be coordinated in partnership with faculty over a semester. The annotated resource set will be supported by UH CI professionals and CI Fellows and CI-TRACS will maintain curricular images/containers and respond to support tickets from researchers, including familiarization, activity dry runs and classroom support.

C. Scalability and sustainability. *Scalability***.** CI-TRACS will create the necessary networks, learning resources, content and programmatic structure to scale the effort in the medium and long terms: (1) medium term scale up of participant numbers to reach wide level of engagement with CLS community in Hawaii; (2) eventual scale out to replicate the model for CI infrastructure training in other domain sciences that have critical need and in which Hawaii has a significant academic research footprint (e.g., astronomy, microbiome science). We note also that CI-TRACS materials are likely to be of interest to industry and business partners. The workshop repository and course container embedding models are designed to be maintained in a CI-TRACS Github repository in HI-DSI. Selected course materials, training materials and workshops will be deposited for a global audience within the Software Carpentry Community Lessons Incubator repository, which will preserve them longitudinally and scale up their accessibility. CI-TRACS may also fill current gaps in the climate science Carpentry portfolio by contributing content on Pacific-focused climate science and areas such as indigenous data sovereignty. *Sustaining impact.* The CI-TRACS will develop a personal and professional network, a community of practice, where cohorts of peers will collaborate and support one another in the long term. Sustained impact and maintaining currency in the field will be supported through integrating participants into the larger existing CI community that is the coordination network around our backbone organization (Campus Champions [CHAMP], Gateway Ambassadors [SGCI], US-Research Software Engineers [US-RSE], CARCC, XSEDE ECSS [XSEDE-ECSS]], see Collective Impact strategy). HI-DSI is being issued with a unique course code identifier to allow cross-listing of all curriculum resources developed in HI-DSI and CI-TRACS and long-term entry of these resources into the UH catalog where applicable. *Sustaining resources.* HI-DSI is an institutional investment by UH System and its longevity in terms of personnel, a data/instructional materials repository and nexus for funding applications will support longitudinal continuation of CI-TRACS. CI-TRACS is the first concerted CI training effort of its kind for Hawaii, and it is our intention for this effort to be a model and launch point for future efforts. As head of CI in the UH System, and heads of HI-DSI, Jacobs and Leigh are committed to developing future grants to NSF and agencies such as DoE and DoD, to help institutionalize CI-training in UH's science and engineering curriculum. We recognize that cutting-edge CI tools designed for scientists are specialized instruments with niche user communities that require sustained efforts for support by funding agencies as companies find them difficult to build a financial model around. Maintaining the necessary CI has required major institutional investment and grant awards, and we intend to supplement this model through industry relationships as CI-TRACS develops. Co-PI Leigh will lead these efforts as he has received numerous awards from DoD (SBIR's from Army to apply SAGE to their workflows), DoE (grants in partnership with the Hawaii State Energy Office to develop advanced visualization capabilities based on the CyberCANOE) and industrial partners (Oceanit, Hawaii Electric Company).

D. Recruitment and Evaluation

D.1. Recruitment process for Climate Fellows. *Target disciplines.* PhD students in climate science, computer science and data science with goal of balance between computer and data science with and atmospheric science, geosciences and ocean science students. *Requirements*: Second year students who have an assigned research supervisor. Good academic standing. *Weighting*: research topic, demonstrated interest in CI, GPA. *Methods*: Flyers, email, social media, website; outreach to Chairs and research faculty. *Application Timeline: Jan 10.* Advertise and promote. *April 1.* Application deadline. *April 15.* Notification. *April 22.* Deadline for acceptance. *May 1.* Final participant confirmed. August 1. Orientation. *June 6-July 1*. Four week UG mentoring in CITRUS. *Goal***:** 4 students per year.

D.2. UG Recruitment process for CITRUS. *Target majors:* climate related natural and social science majors. *Criteria:* at least sophomore standing, good academic standing. *Weighting:* prior interest/research experiences, stated interest in CI, GPA. *Sources*: 4 year institutions in Hawaii and US-affiliated Pacific region (leveraging the NSF Pacific LSAMP network of 11 institutions, NSF HRD1826864, Turner, co-PI). *Methods*: Flyers, email, social media and website, using LSAMP coordinators as POC. This worked successfully for example for the Pacific Datacamp program at CUH. *Application Timeline: Feb 1.* Advertise and promote participation. *April 1.* Deadline for applications. *April 15.* Notification. *April 22.* Deadline for acceptance. *May 1.* Final participant roster confirmed, special interest groups assigned *May 23-June3.* Two week pre-CITRUS bootcamp using training modules. *June 6th.* Orientation. *June 6-July 1*. Four week immersion experience. *July 1.* Capstone showcase presentation. *Notification.* Participants will be notified with terms and conditions and a post-acceptance survey will determine (a) their tech needs and environment, and (b) their preference for CLS focus areas (e.g., atmospheric science, geosciences, ocean science). *Goal***:** 12 students per year.

D.3. Diversity recruitment strategy. DEI lead Turner will coordinate pipeline efforts to ensure a diverse cohort with female and indigenous student representation for CITRUS and CI Fellows. We will engage NHPI and female role models in the recruitment process (faculty, HI-DSI Data Science fellows). Female role models and peers will be engaged to assist with recruitment, and campus organizations/clubs such as Graduate Women in Science Hawaii [GWISH] will be targeted for recruitment and outreach. The NH-serving UH Hilo (ICS and Environmental Science programs, and CUH (Data Science and Environmental Science programs) will be targeted with recruitment materials and presentations both to recruit for CITRUS and form a pipeline into UHM graduate programs that would be eligible to be eventual CI Fellows. Recruitment for graduate student participation as CI Fellows will focus on students who are at least 1 year into their graduate degree. NHPI serving campus units (Center for Pacific Island Students at UHH, Native Hawaiian Student Success program at UHM, SACNAS Hawaii Ilima Chapter (UHM, CUH, UHH), Ho'oulu STEM and HHMI Inclusive Excellence programs at CUH) will be made aware of the CITRUS and CI Fellows opportunities and enlisted for recruitment. Culturally-normed informational sessions (online or in person) will include near peer students, information on the program that can be readily shared with family and community, especially Kupuna (elders), and a focus on the long term opportunity to be part of a community/affinity group (hui). Explicit recruitment and inclusivity strategies for gender minorities and non-binary student outreach will be coordinated by CI-TRACS team member Stokes, who is a member of the UHM Commission on LGBTQ+. Across CITRUS and CI Fellows, inclusivity will be also fostered through intentional focus on the applicability of CI skills to regional challenges in climate science that are of key importance to families and communities in the Pacific. ***Inclusivity.*** CI-TRACS will promote bias reduction. All participants will train on reduction of bias and systemic racism. Specific programming on areas such as algorithmic bias will be included in the CI-TRACS curricula. Inclusion and sense of belonging will be promoted through programming (seminars, guest speakers), CyberClub activities (film screenings, guided discussions), culturally-relevant content (data sovereignty, indigenous climate knowledge) and culturally-normed opportunities for community engagement and focus on regional problem sets.

D.4. Evaluation. ***Independent evaluation.*** Dr. Suzanne Pierce will serve as independent evaluator. She is a Research Scientist with TACC and Environmental Science Institute in the Jackson School of

Geosciences. Her community based research uses intelligent systems to understand Earth, including water, energy, urbanization, and ecosystem services. Dr. Pierce will periodically review and evaluate outcomes from the program and report annually to an all-hands meeting (see Management and Coordination Plan). The PI, co-PIs and support staff will lead the project tracking program, and deploy instruments such as surveys, to inform Dr. Pierce's evaluation. External evaluation will focus on *formative* information to guide program improvement and *summative* assessment of program quality, effectiveness, and impact. CI-TRACS evaluation will answer 4 questions: *(1) Implementation*: Are activities being implemented on schedule as planned? *Instruments:* (a) review of timeline and project tracking information, reporting to all-hands meeting with remediation/mitigation plans for any missed milestones; (b) database of participant numbers and basic demographics maintained by MT. *(2) Effectiveness:* Are key components of CI-TRACS (student recruitment, selection, mentoring; curriculum; research activities) operating effectively and how might they be improved? *Instruments:* Student survey and reflection instruments. Quarterly reports by activity leaders, comprising metrics relative to benchmarking, assessment of accomplishment, barriers and challenges, development mitigation plan for challenges. *(3) Impact:* What near (scholar demographics, performance, CI attitudes) and longer-term outcomes (academic and employment/graduate placement) are associated with CI-TRACS participation? What is CI-TRACS effect upon CI capacity in CLS UH research? *Instruments*. Participant pre- and post-surveys, student/faculty reflections. *(4) Sustainability:* How and to what extent are elements of the CI-TRACS program scaling and becoming institutionalized? *Instruments:* Leadership/ stakeholder surveys and reflections.

D.5. Advisors. We have identified internal and external advisors to assist CI-TRACS. *Climate Scientists:* Tom Giambelluca, Co-I, John Burns (UHH, Coral reefs), Guiseppe Torri (UHM, Climate modeling/prediction). *ICS faculty.* Peter Sadowksi (UHM, AI/ML, curriculum), Mahdi Belcaid (UHM, informatics, curriculum); Henri Casanova (UHM, HPC); *Diversity and Recruitment:* Cameron Miyamoto (UHM, Director LGBTQ+ Center, UHM); Laura Tipton (CUH/HI-DSI, Women in Data Science and Graduate Women in Science). Joseph Genz, Marata Tamaira (UHH, LSAMP and UHH Center for Pacific Island Students liaison). **External Advisory Board**: Chris North (Virginia Tech, data science), Maytal Dahan (TACC, systems science), David Tarboton (Utah SU - Hydroshare, Hydroserver, FAIR & monitoring data), Carol Song (Purdue - GeoEDF Geospatial & FAIR Data), Parker Antin (CYVERSE), Cathy Constable (Scripps, Earthcube).

E. Collective Impact Strategy.

The CI-TRACS Cybertraining program is positioned for collective impact on Cyberinfrastructure training in Hawaii with eventual scaling to the US-affiliated Pacific region. The CI-TRACS program responds to elements of a Collective Impact Framework as shown below, with an emphasis on the formation of a sustainable Backbone Organization, and continuous communication to advance a mutually-shared agenda on empowering the future climate science workforce serving the Pacific with advanced CI skills and competencies. ***Collective Impact Strategy 1:* Form a Backbone organization and coordination network.** The Backbone organization is the inter-institutional partnership between UHM, UHH and CUH. Within these campuses the organizational units of HI-DSI, UH-ITS, academic departments with climate and computer sciences and data science focuses, and funded programs (EPSCoR, HHMI, INCLUDES, TAPIS, SAGE) form an internal network. A wider coordination network comprises *partnerships* (with TACC, PEARC, CARCC, Science Gateways, IGP Viz, CYBERHawaii) and *resources* (Datacamp, Software Carpentries, Jetstream 2). ***Collective Impact Strategy 2:* Pursue a Shared Agenda.** The mission and vision of the participants in the backbone organization and coordination network are aligned and reflect shared agendas to improve CI education, empower research through CI, develop workforce to meet meets and create opportunity and to ensure that CI resources and careers are available equitably and inclusively to all communities. This shared agenda will be developed in phases commonly associated with a Collective Impact approach. *Phase 1. Ideation and Dialogue*. Approximately 7 years of dialogue, collaborative projects, shared programs between the backbone organizations and coordination network have laid foundations for CI-TRACS, a pilot phase which leads to the current implementation level proposal. *Phase 2. Initiate action*. CI-TRACS emerges at a point in time where the backbone organizations and coordination partners have initiated a number of shared actions to address critical gaps in programming and academic opportunity, with a current focus on CI. *Phase 3. Organize for impact.* CI-TRACS' Collective Impact strategy will formally organize for the first time the partnerships and connections that make up the backbone organization and coordination networks. _Phase 4. Sustain action and impact._ CI-TRACS represents a first intentional approach to build CI capacity in a specific domain science (climate science). Sustained impact can be realized by replicating and expanding this model to other domain areas. **_Collective Impact Strategy 3: Ensure Continuous Communication._** The communication and dissemination capacity built in CI-TRACS will enable new information sharing across the coordination network. The MT and Advisory Board include coordination network personnel, and the dissemination of resources and human connectivity (multiple CI-TRACS players have multiple roles across the network) will enable collective understanding and **Mutual Reinforcement** **_(Collective Impact Strategy 3)_** of diverse efforts in CI across organizations. CI-TRACS Collective Impact strategy will ensure that the program's Identified **Measurable Outcomes** (**_Collective Impact Strategy 4)_** are keyed to wider national goals and imperatives for CI training.

Moreover, through the coordination network, CI-TRACS becomes part of the wider conversation concerning desired measurable outcomes in a national effort to increase CI capacity in support of academic research, workforce development and educational opportunity.

F. Fostering Community. CI-TRACS will foster community at three levels. First, the coordination network described in our Collective Impacts framework is a broad community of practice for CI professionals, users and contributors. CI-TRACS will serve a virtual convening function for conversations within this community. We will foster awareness and engagement of our participants in organizations such as those listed in A.2., e.g., by publicizing and supporting opportunities such as professional memberships, workshop attendance and conferences. Second, climate scientists who will be participating to strengthen their CI skills as part of their own community of practice through connection to efforts such as Earthcube, Science Gateways and special interest groups at the American Geophysical Union (AGU). Third, participants and trainees will be fostered as a collaborative learning community through CyberClub, which will run a Discord/Slack space for participants and community to meet, support one another and engage in shared problem solving, encourage Vlogs and Podcasts, and develop events such as an annual climate data hackathon.

G. Broader Impacts.

Advance discovery and understanding while promoting teaching, training, and learning. The research workforce in CLS will be strengthened through 36 undergraduate and 12 graduate

training experiences in CI. The availability of 10 new workshops and embeddable curriculum modules will increase CI awareness and skills across the CLS domain in research and education at UH with resultant positive impacts on CLS research sophistication and productivity. ***Broaden participation of under-represented groups.*** UH and CUH are both minority-serving, indigenous serving institutions with high proportion of low income and first generation students. Intentional inclusion of diverse and indigenous trainees will broaden participation and increase career competitiveness needed to promote self-determination of the climate future for students from Pacific island communities. ***Broaden dissemination to enhance scientific and technological understanding.*** The Collective Impact framework and coordination network provides for broad dissemination, and specific efforts such as inclusion of CI-TRACS resources into repositories such as Software Carpentry will support broad adoption. ***Benefits to society.*** Climate Science is critical to the Pacific future. UH has built research excellence and a Hawaii-Pacific

H. Timeline and Workplan	*Year 1*	FA	SP	SU	**2**	FA	SP	SU	**3**	FA	SP	SU
Broaden adoption of CI skills												
1A CI Fellows	*Recruit, orient*											
	CI Training Workshops											
	Capstone presentation											
	CyberClub											
1B CITRUS	*Recruit, select*											
	2wk skills,4wk research											
	Capstone presentation											
	CyberClub											
CI Integration into Curricula												
2A: Workshop development												
2B Course modules	*RFP to faculty*											
	Course redesign											
	Deploy, evaluate, refine											
Mgmt. Plan	*MT meetings weekly*											
	Monthly all hands											
	Annual kickoff											
	EAB/External evaluator											
	Coordination network											

workforce pipeline in climate science that it now seeks to empower with critical cyberinfrastructure skills. Enhanced academic climate science at UH will benefit society through informing Pacific responses to climate change and increasing opportunity for Pacific communities to determine their climate futures through full participation in the discipline.

I. Results from Prior Support.

Gwen Jacobs, PI. (1) OIA-1557349, $20M 06/01/16-05/31/21. RII Track-1: Ike Wai*: Securing Hawaii's Water Future* **Intellectual Merit**: Ike Wai has built capacity for hydrogeological research, created workforce pathways for research and industry, and produced actionable information for water managers and policy makers in Hawaii . Isotope measurements in submarine groundwater discharge (SGD) and geochemical analyses of 400 well and pond samples suggest a lack of aquifer boundaries in highly porous basalts. A marine geophysics study revealed deep reservoirs of low salinity submarine groundwater suggesting a new land-to-sea groundwater transport mechanism. The Ike Wai Science Gateway supports data management, computation, analysis, visualization, and dissemination of all data and data products. **Broader Impacts**: 1375 individuals have participated in Ike Wai including 33 faculty, 5 postdoctoral fellows, 24 graduate students, six non-technical support staff, 13 technical support staff, and 54 UG researchers. Completed 14 graduate degrees, 18 UG degrees. Five new faculty at UHM (geology, engineering, economics and geophysics). At UHH 4 new faculty hired in Data Science. **Products:** 89 peer-reviewed publications [IkeWaiPubs], 122 proposals submitted, 9 NSF awards ($2,870,025) and 1 CAREER award (Mandel, UHH). **(2)** NSF ACI-SI2 #1450413 "Collaborative Research SS2-SSI: The Agave Platform: An Open Science-As-A-Service Cloud Platform For Reproducible Science**"** $3,899,909, 08/01/2015-07/31/2019 **Intellectual Merit:** This project supported the extension and improvement of the Agave API Cloud platform features and the deployment of a science gateway to support the Hawaii EPSCoR Ike Wai project. The gateway provides data management, computation, analysis, and visualization tools to support modeling, data discovery, and decision support. **Broader Impacts***:* This project broadens the range of users to students and community colleges use advanced CI resources through web interfaces that manage the complexity of computational workflows, access to remote computing resources, data transfer. Pubs [Cleveland18a,b].

Jason Leigh, PI OAC-1441963, $5,204,441, 05/14-09/20. *SI2-SSI: SAGEnext: Next Generation Integrated Persistent Visualization and Collaboration Services for Global Cyberinfrastructure***. Intellectual Merit:** SAGE and its successor SAGE2 are middleware platforms that enable groups of users to effectively use ultra-resolution collaboration environments. **Broader Impacts:** Since 2004, ~800 institutions have used SAGE1&2 to enable collaborative scientific data visualization. **Products:** 27 publications as a result of SAGE and SAGE2 efforts [SagePubs]. The software has been openly distributed and continually improved since 2004. A 2020 award (OAC-2004014) from NSF continues the effort into SAGE3 that integrates AI.

Helen Turner (1) NSF INCLUDES #1744526, 09/15/17 to 08/31/20. PI Gaither, co-PI Turner. $299,996. Title. *NSF INCLUDES DDLP: SPICE (Supporting Pacific Indigenous Computing Excellence) Data Science Program for Native Hawaiians and Pacific Islanders*. **Results.** Intellectual Merit. SPICE established a model for DS preparation of Native Hawaiian Pacific Islander (NHPI) students that is contemporary, culturally-consistent and sustainable. SPICE developed the Pacific region's first BS in DS, and built capacity to use DS for community and social impact in the Pacific. Broader Impacts. SPICE has provided innovative educational approaches to prepare the next generation of Hawaiian and Pacific Islander scientists to lead the development of data-enabled solutions to Pacific regional challenges. **Publications.** 2 to date**. Evidence of research products and their availability.** Projects are found on Chaminade's website [INCLUDES]. **Turner PI. (2) NSF I-USE DUE1525884, 10/01/15 to 09/30/18 CUH: $300,000 total costs. Title.** *Kûlia: CUH University I-USE Project*. **Results.** Intellectual Merit. Kûlia is an inter-institutional effort (CUH University of Honolulu and the University of Hawaii at Hilo) that focuses on "big data" and the curricular infusion of indigenous knowledge. Broader Impacts. Kûlia has provided innovative educational approaches to prepare the next generation of Hawaiian and Pacific Islander scientists to lead the development of practical and policy-based solutions to Pacific regional challenges. **Publications.** [Chong 2019], [Gaither, 2018]. **Evidence of research products and their availability.** Curriculum modules are made available on the Chaminade website [IUSE] **(3) Turner PI, Stokes co-PI. NSF S-STEM DUE 2030654, 10/01/20 to 09/30/25. $1M. Title:** Pacific Excellence in Analytics through Research and Learning in Data Science*.* **Results.** *Intellectual Merit.* Established the *PEARL* Pacific Scholars S-STEM scholarship program for 20 undergraduate Data Science majors at Chaminade with first incoming cohort Fall 2021. **Publications**. None to date. **Evidence of research products/availability.** None to date.