Research of Impact:

CRC Annual Report 2022-23

University of Pittsburgh®

Table of Contents

3 CRC Users Lead Research at Pitt

4 Cyberinfrastructure Upgrades

5 AWS Cloud Seed Program

5 Faculty Recruiting Leveraging CRC Resources

6 ARC23 Symposium: COVID-19 Wins and Lessons

7 Training and Workshops Drive CRC Outreach Mission

8 User Statistics Reveal Growth, Enhanced Efficiencies

9 Cluster Utilization and Top 10 Departments with the Most Core-hours

10 Buildout of SOP and Internal Standards

11 Building Team Expertise

CRC Users Lead Research at Pitt

The Center for Research Computing can report that we supported an unprecedented level of funded research leading to prominent publications in FY23. That is not a new situation: we have known for years that CRC resources contribute to a vast range of work within Pitt's research enterprise. During the past five years, CRC has been steadily improving our data collection and documentation of the funding levels of projects we support:

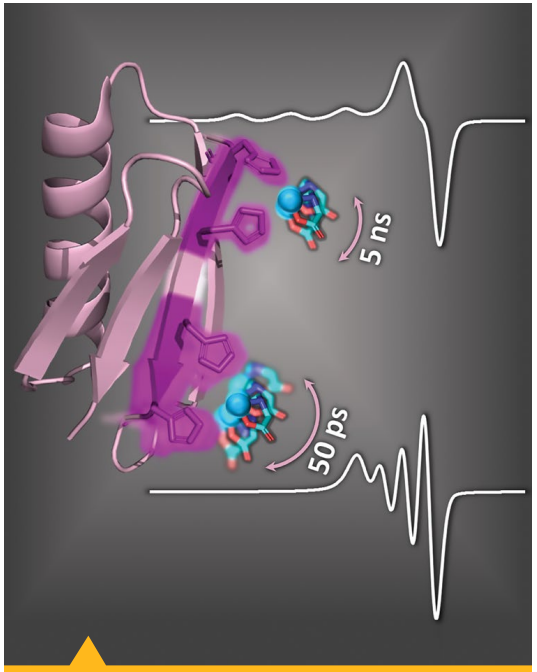
254 GRANTS | **$342M** EXTERNAL FUNDING

In FY23, CRC enabled over 254 active grants, supported by over $342,000,000 in external funding. CRC collaborators published more than 189 papers and book chapters.

- **2017-2018: CRC enabled more than 116 active grants totaling over $64 million in external funding.**
- **2019-2020: CRC enabled 184 active grants supported by $143 million in external funding.**
- **2020-2021: CRC enabled over 154 active grants supported by $124.4 million in external funding.**
- **2021-2022: CRC enabled 155 active grants supported by almost $160 million in external funding.**

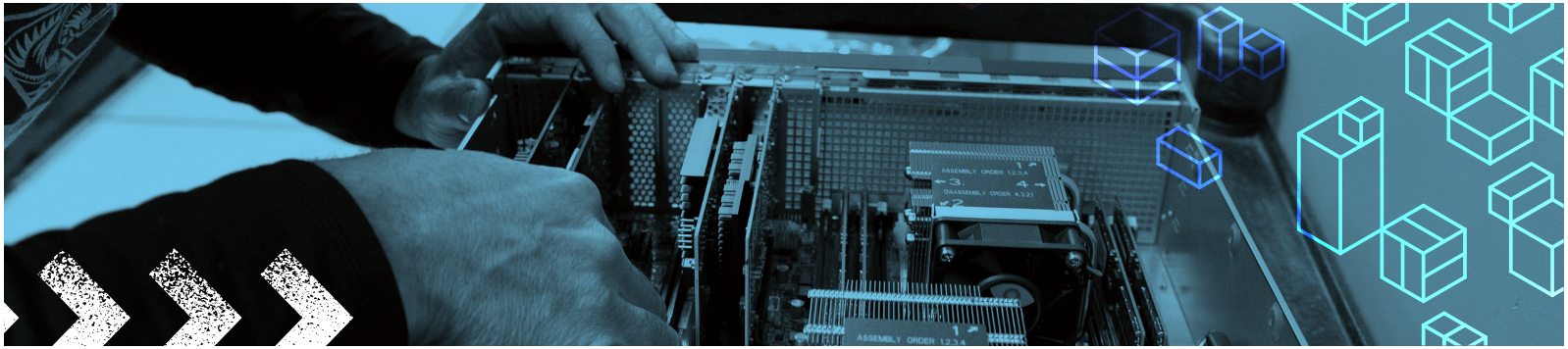
Our processes have greatly improved, and in FY23, CRC enabled over **254 active grants**, supported by over **$342 million in external funding***. With research at Pitt surpassing a landmark of $1 billion in research expenditures, CRC enables and consults on projects representing nearly one-third of the university's external funding. CRC plays a significant role in helping Pitt solidify its place among an elite national cohort of research peers.

*These are the grants for all PIs that used our clusters in FY23 (July 1, 2022-June 30, 2023). Some proposals were submitted in FY22. These PIs used the clusters both in FY22 and in FY23. Some proposals submitted in FY23 used the clusters this year and will continue to use it next fiscal year (FY24).



Enabling Advanced Molecular Dynamic Simulation

This molecular dynamic simulation was developed by the lab of Pitt chemistry professor and Chair of the department of chemistry Sunil Saxena. It illustrates two sites of a protein with different site-specific dynamics (shown in purple), measured by a newly developed methodology that helps elucidate the role of protein dynamics to protein function. The image was developed using CRC resources and was the cover feature for the October 2022 issue of the European Chemical Societies Publishing journal *Analysis & Sensing*.



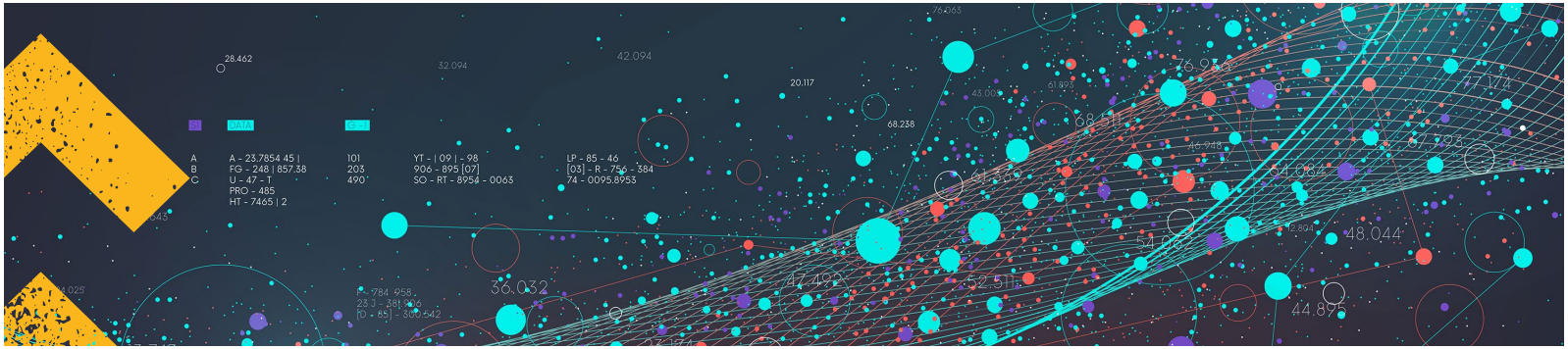
Cyberinfrastructure Upgrades

Hardware
The major hardware upgrade for FY23 is a refresh of the SMP cluster. We executed an RFQ and ordered 53 compute nodes with 64 cores within each to replace 132 nodes (24 cores/node). The new hardware will have more performant memory (12GB/core of DDR5 vs. 8GB/core of DDR4) to support the ever-increasing higher-fidelity models and data sciences workflows. The new compute nodes will provide an estimated aggregate performance of 280 Teraflops compared to the 130 Teraflops that were replaced in the retired nodes.

Software
CRC partnered with Swanson School of Engineering and Kenneth P. Dietrich School of Arts and Sciences to purchase an effective institution-wide pool of COMSOL Multiphysics licenses. CRC purchased a pool of licenses for broadly used software on the CRC ecosystem, including MATLAB, Stata, SAS and Mathematica.

Protected Environment
CRC prototyped a HIPAA-compliant computing and data environment that is suitable for storing and analyzing personal health information. The environment consists of 600TB of usable storage built from self-encrypting drives, a login node, a GUI portal node, a cluster of repurposed CPU nodes (quantity 16), and a cluster of repurposed gaming GPU nodes (quantity 4). We plan to onboard early users from the first project in June, and then work with additional users from two separate projects in July.

Dell Moonshot
The Dell Moonshot project is a collaboration between the University of Pittsburgh, University of Pittsburgh Medical Center, Pittsburgh Supercomputing Center (PSC), Carnegie Mellon University and Dell to create an Innovation Hub for Health Science Medical Research to bring personalized medicine to the masses through different biomedical science focus areas and diverse patient pipelines. It was catalyzed through an in-kind donation of FMV approximately $1.6 million in infrastructure and services from Dell and augmented through additional infrastructure from Pitt IT. The equipment is being installed at the Pitt data center. The broad team from Pitt IT, CRC, PSC, and Pitt's Department of Biomedical Informatics have engaged in meetings to map the various project requirements to the hardware, networking, and security infrastructure, with the goal of providing the developers access to the environment by early July.



aws Cloud Seed Program

CRC is collaborating with Pitt IT and Amazon Web Services (AWS) in a program that offers up to $5,000 funding in cloud computing credits to allow investigators to explore the application of AWS Cloud Computing to their research projects. As of May 30, 2023, **30 proposals** have been received with **16 approved**. Applications for FY23 have been closed.

The program demonstrates to researchers the great benefits offered by Cloud Computing – virtually unlimited resources, no queues, innovative applications, no hardware to buy or manage and the freedom to compute from any device, anywhere.

Researchers wanting to use cloud computing typically run up against barriers of funding, training and support. In a joint effort to remove those barriers and help ease Pitt researchers seamlessly into the cloud ecosystem, the AWS Cloud Seed Program aims to create a sustainable relationship where researchers can build a robust cloud architecture that augments existing research workflows.

The program offers up to $5,000 funding for AWS cloud credits for appropriate research projects, offering:

- **Consultation and technical support from domain experts**
- **On-site training and virtual workshops**
- **Secure collaborative data environments supporting compliance requirements**
- **Engagement with other researchers leveraging AWS**

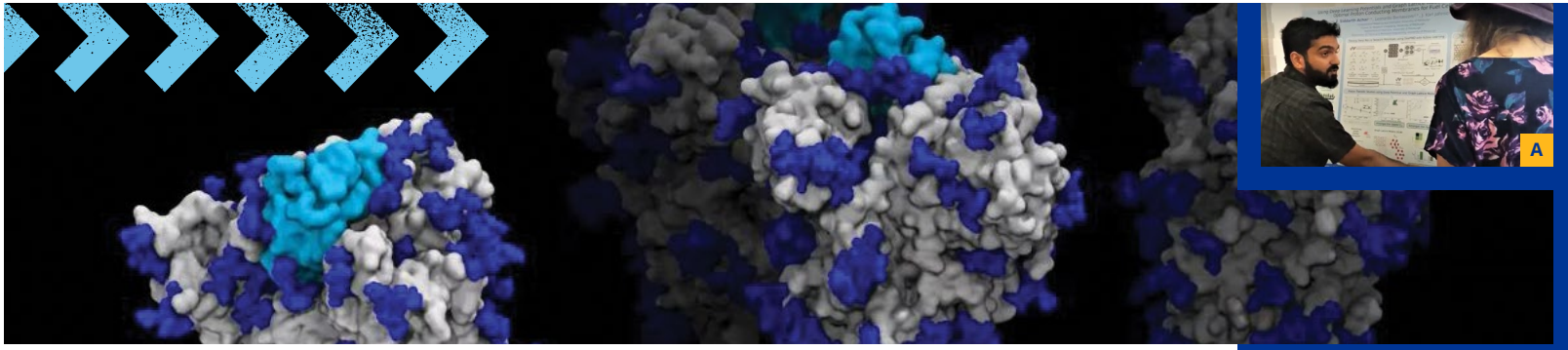
Faculty Recruiting Leveraging CRC Resources

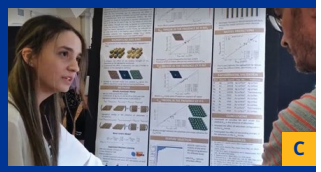
Biological Sciences
Cara Haney, University of British Columbia.
microbiology.ubc.ca/research/labs/haney | Jan 31, 2023

Nathan Clark, University of Utah.
clark.genetics.utah.edu | Jan 12, 2023

Physics
Zhehao Dai, University of California, Berkeley.
scholar.google.com/citations?user=VfZfMm0AAAAJ&hl=en
April 19, 2023

Chemistry
Grace Kenney, Harvard.
chemistry.harvard.edu/people/Grace-Kenney
March 28, 2023





ARC23 Symposium: COVID-19 Wins and Lessons

Rommie Amaro

Lillian Chong

In April, CRC hosted leading researchers in fields from molecular dynamics to public health at the Center's 2023 Advancing Research through Computing Symposium, ARC23. The theme was COVID-19 Wins and Lessons, and speakers explored the ways that advanced computing was employed in research into the pandemic.

Following the theme, the symposium opened with two back-to-back speakers who worked on one of the great "wins." Keynote speaker **Rommie Amaro**, distinguished professor of theoretical and computational chemistry at University of California San Diego, and Pitt chemistry professor **Lillian Chong** both worked on one of the most successful and eye-opening discoveries into the functioning of the SARS-CoV-2 virus. The team developed a molecular dynamics simulation of the SARS-CoV-2 virus in atomic detail to show the function of the spike protein as the virus infects a human cell, relying on the WESTPA sampling algorithm Chong developed with CRC resources.

The team's study, published in Nature Chemistry, advanced understanding of how the opening of the spike protein – the spiky structures protruding from the round surface of the cell of the virus infect the host cell by one glycan – sugar molecules covering the spike protein – acts as a lever, pushing the spike receptor from a "down" to an "up" position. Collaborators from two experimental labs validated the results of the simulations that were published in The New York Times. In recognition of this work, the team won the 2020 ACM Gordon Bell Special Prize for HPC-Based COVID-19 Research, widely known as the Nobel Prize of Supercomputing.

At the ARC23 poster contest, 18 Pitt graduate and undergraduate students competing for one of three $750 travel scholarships, presenting an impressive array of work from across schools and majors.

The winners: Siddarth Achar (A), Chemical Engineering, "Using Deep Learning Potentials and Graph Lattice Models to Engineer Optimal Proton Conducing Membranes for Fuel Cells"; **Meiirbek Islamov (B)**, Chemical and Petroleum Engineering, "A Data-Driven Exploration of Structure-Property Relationships of Thermal Transport in Metal-Organic Frameworks"; **Maya Salem (C)**, Chemical Engineering, "Understanding the Segregation Energy Behavior of Single Atom Alloys in the Presence of Ligands"

Training and Workshops Drive CRC Outreach Mission

The CRC team presented and hosted 38 workshops in FY23, with the core team presenting five in Fall 2022 and six in Spring 2023 that encompassed a wide range of data science, offering introductory tutorials covering general programming in Python, object-oriented programming, and data handling and visualization. Each workshop had about 150 participants, both those who attended the live workshop and those who viewed the recorded workshops asynchronously. Additionally, the CRC organized a series of hands-on workshops focused on industry-level topics. These workshops explored accelerated data handling, manipulation, and visualization techniques, as well as accelerated machine learning using Nvidia graphics processing units (GPUs). They also delved into advanced deep learning methods, such as convolutional neural networks, transformers, and language models, along with training massive-data dependent deep learning models on Nvidia GPUs. Each of these workshops had an attendance of 120 participants.

Fangping Mu

Fangping Mu, research assistant professor and consultant at CRC, continued the successful series of Next Generation Sequencing (NGS) Workshops funded by a 2021 Pitt Seed Grant, including five two-hour workshops in Summer 2022 with 40-50 attendees per workshop. The NGS series presented 20 three-hour NGS workshops in Fall 2022 and Spring 2023, each with 10-30 participants.

CRC's data science workshops were attended by a total of **150** registered participants. Participants were largely Pitt (including Pitt-Johnstown) students, postdocs, and faculty, but there were also attendants from Howard University, University of Texas Health at San Antonio and The University of Memphis.

CRC co-hosted a 5-day hybrid workshop with the MEMS on "Python for Scientific Computing and TensorFlow for AI", presented by Manchester Metropolitan University national teaching fellow Stephen Lynch. The Python workshop attracted over **60 in-person** and **25 online attendees**.

CRC consultants worked with about **150 students and researchers** through one-on-one consultations.

CRC also provides support for formal classes:

ME 2054 Parallel Computing for Engineers taught by Inanc Senocak, associate professor in Mechanical and Materials Science (on 4 GPU nodes)

ME2060 Numerical Methods taught jointly by Senocak and Engineering HPC consultant Chengnian Xiao (with Julia 1.6.4)

User Statistics Reveal Growth, Enhanced Efficiencies

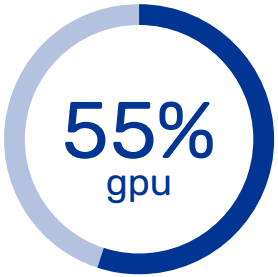
Highlights of usage statistics from Jan. 1, 2022 to April 30, 2023						
	All Users	**Faculty**	**Grad/ Post-doc**	**Sponsored**	**Staff**	**Undergrad**
Users	1192	138	652	112	109	223
Departments	105	57	83	30	38	42
Core Hours	>109M	>40M	>58M	>4M	>2M	>3M

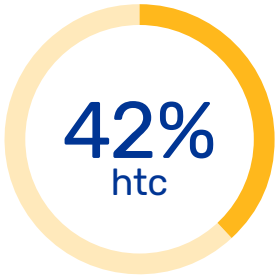
The CRC remains committed to increasing transparency, ensuring Pitt's community and users have access to comprehensive usage statistics. As a testament to this commitment, the CRC unveiled an immersive usage statistics dashboard on its homepage last year (https://crc.pitt.edu/stats). This powerful tool empowers users to explore the evolving utilization of clusters over time, gaining valuable insights into how resources are distributed among various users and departments. In an effort to continuously improve, the dashboard keeps evolving, incorporating even more helpful information to aid both prospective and current users. For instance, average wait times on each cluster will soon be available allowing users to make well-informed decisions about resource allocation and workflow management when engaging with the CRC. CRC expanded its provision of resources to accommodate a remarkable **8 percent increase in users** spanning across **105 departments** in comparison to the period Jan. 1, 2021, to May 31, 2022, which resulted in more than **109 million core-hours** on all clusters.

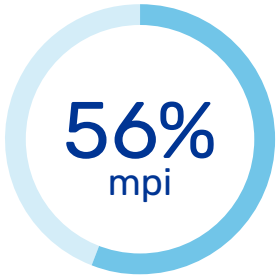
Cluster	Partition	Average Wait Time (hh:mm:ss)	Median Wait Time (hh:mm:ss)
GPU	A100 A100 Multi A100 NVLINK	00:11:29 00:31:12 00:15:29	00:00:02 00:00:03 00:00:01
HTC	HTC	00:47:07	00:00:02
MPI	MPI OPA HIGH MEM	00:18:19 00:00:06	00:00:01 00:00:06
SMP	HIGH MEM SMP	00:19:30 00:02:58	00:00:01 00:00:17

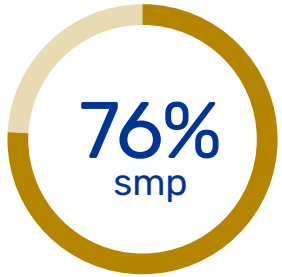
This table displays the average and median wait times from when a job is submitted to when it starts running on all clusters and partitions of the CRC. The data reveals that the majority of jobs submitted to the CRC clusters commence within minutes, even in the case of high-demand clusters/partitions like the multiple GPU partitions and high-memory SMP cluster. These brief wait times exemplify the efficiency and smooth integration of hardware and software platforms, fostering a compelling environment that attracts more users to apply for and make use of the CRC's valuable resources.

Cluster Utilization and Top 10 Departments with the Most Core-hours



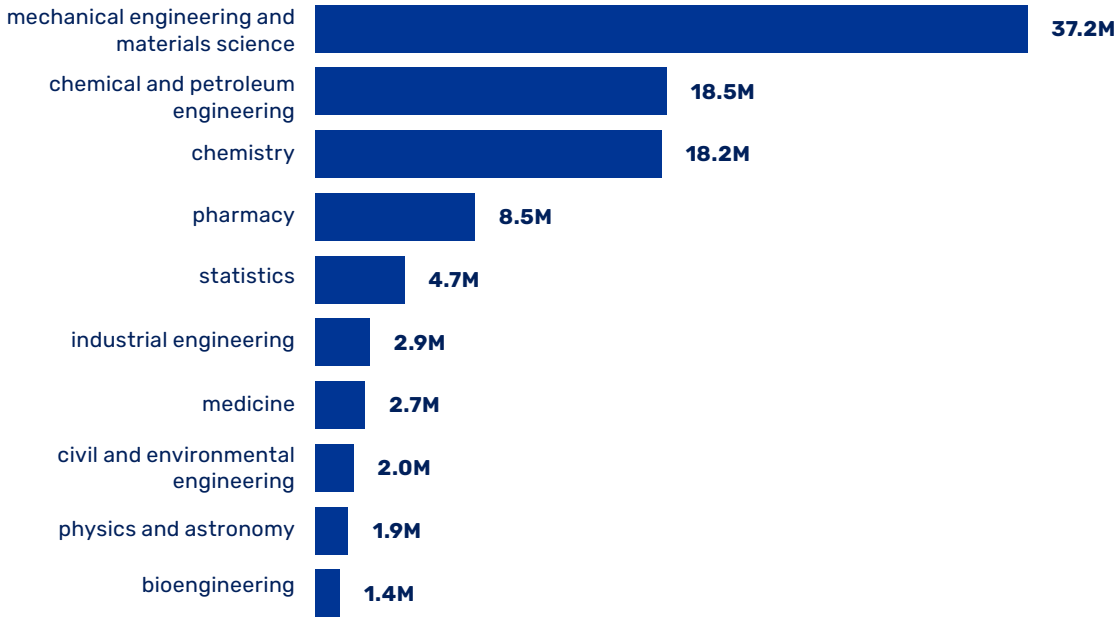






Responding to User Help Tickets

Between Jan. 1, 2022, and April 30, 2023, the CRC team members offered assistance and guidance in over **3,300 tickets** and requests. These tickets covered a wide range of issues, from basic tasks like account (storage and computing) administration and software installations to specialized support in domains such as health informatics, computational biology and data science, specifically related to running jobs and pipelines on the cluster. To better facilitate users' work on clusters and resource management, the CRC also created various tutorials based on frequently requested features and issues, making them accessible to all users.



(out of a total **105** departments)

Buildout of SOP and Internal Standards

Internal Standards
CRC has migrated all software development activities onto the University's GitHub enterprise platform. This has resulted in significantly improved collaboration between team members by promoting team-based development workflows.

Software deployment
The adoption of GitHub as a version control utility, combined with the implementation of new operational infrastructure, has streamlined the CRC's ability to deploy and rollback software changes. This has resulted in significantly faster software deployments and reduced response times to software issues that may otherwise impact the experience for researchers and students.

Advanced Automation
CRC has improved its operational efficiency by automating a large number of previously manual tasks for software standards. This automation saves time while ensuring work is performed consistently and to a high degree of quality.

Automated software tasks include:

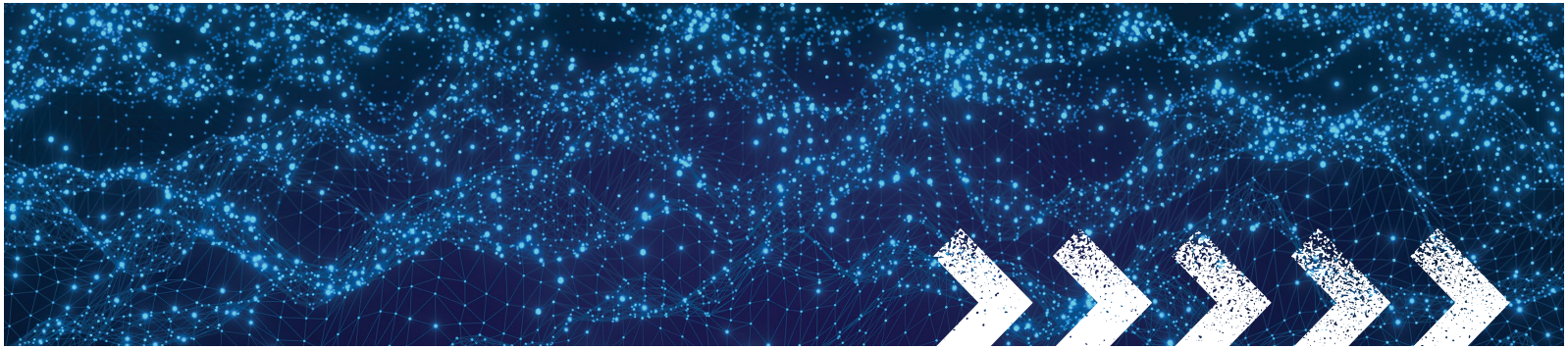
- Testing
- Documentation management/publication
- Publication/delivery
- Quality inspections
- Dependency management
- Security vulnerability scanning

SOP
The CRC is utilizing GitHub to build and host a set of clean and easily navigable internal standard operating procedure (SOP) documents. There are currently 26 documents in our internal SOP, with 13 recommendations from team members for new processes and improvements.

The intention behind maintaining this documentation is to concisely explain common cluster administration tasks so new team members can easily perform them, and so that knowledge is retained as the composition of the team changes over time. Adhering to the established instructions enables the iterative refinement of the document and the process as they are used and changed should someone produce a better solution or automate one or more steps.

Some important policies in place for the development of SOP are:

- **At least two** members must know and have verified the process
- New SOPs require review before they are adopted
- Changes need approval by one other team member
- Topics include:
  - Streamlined instructions for creating SLURM allocations and user accounts
  - Performing software installation
  - Awarding service units
  - Provisioning storage



Building Team Expertise

CRC expanded its available expertise this year with the addition of two team members, one replacing our faculty researcher specializing in engineering applications and one in the new role of data science expert that we have sought to recruit for several years.

Chengnian (Cheng) Xiao

Chengnian (Cheng) Xiao recently joined the CRC team as an engineering HPC consultant. With a master's degree in applied mathematics and a PhD in mechanical engineering, he most recently worked as a postdoc studying complex turbulent flows using both mathematical analysis as well as exascale computing in Pitt's Department of Mechanical Engineering and Materials Science. With several years of experience in theoretical engineering, fluid dynamics, and scientific computing, Cheng is excited to be part of CRC's mission at the intersection of tech and domain research.

Yassin Khalifa

Data scientist **Yassin Khalifa** is the newest addition to the CRC team. With a diverse background in biomedical engineering and computer science, Yassin has made significant contributions to medical research on multiple projects that utilize artificial intelligence to develop assistive technologies for the diagnosis and prevention of diseases. Yassin earned his PhD from the department of Electrical and Computer Engineering at Pitt.

He has worked in Pitt's Innovative Medical Engineering Developments laboratory developing computational approaches that use medical images and signal data to understand the effects of disease and aging. Most recently at Case Western Reserve University, he contributed to research on the impact of the built environment and long-term exposure to air pollution on the development of cardiovascular disorders.

University of Pittsburgh

Center for Research Computing

312 Schenley Place
4420 Bayard Street
Pittsburgh, PA 15260

412-648-3094

crc.pitt.edu

114201-0623