



# PURDUE RESEARCH COMPUTING

## FY 2019 ANNUAL REPORT

Image courtesy of Daisuke Kihara, Purdue professor of biological sciences and computer science, who uses Purdue's supercomputers to model protein structure and interaction.

# DIRECTOR'S WELCOME

It is with great pleasure that I welcome you to this first-ever annual report highlighting Research Computing at Purdue University. Purdue posted record levels of sponsored research awards in 2019 – \$520.6 million – and Research Computing partners accounted for \$262 million of those sponsored dollars – over 50% of all awards!

Research Computing comes from a lineage dating back to the 1960s, when Purdue became one of the first universities with a supercomputer, and deployed a CDC 6500 system. Our organizational ancestors operated for Purdue's scientific community many systems, advanced for their time, and we are proud to carry on this tradition of advanced computing at Purdue.



**Preston Smith,**  
Executive Director,  
Research Computing

The fall of 2019 marks the 15th year of Purdue's nationally recognized community cluster program, begun in 2004 under the leadership of Purdue's first CIO, Jim Bottum. *Hamlet*, the first community cluster, started with just a handful of machines and six faculty partners,

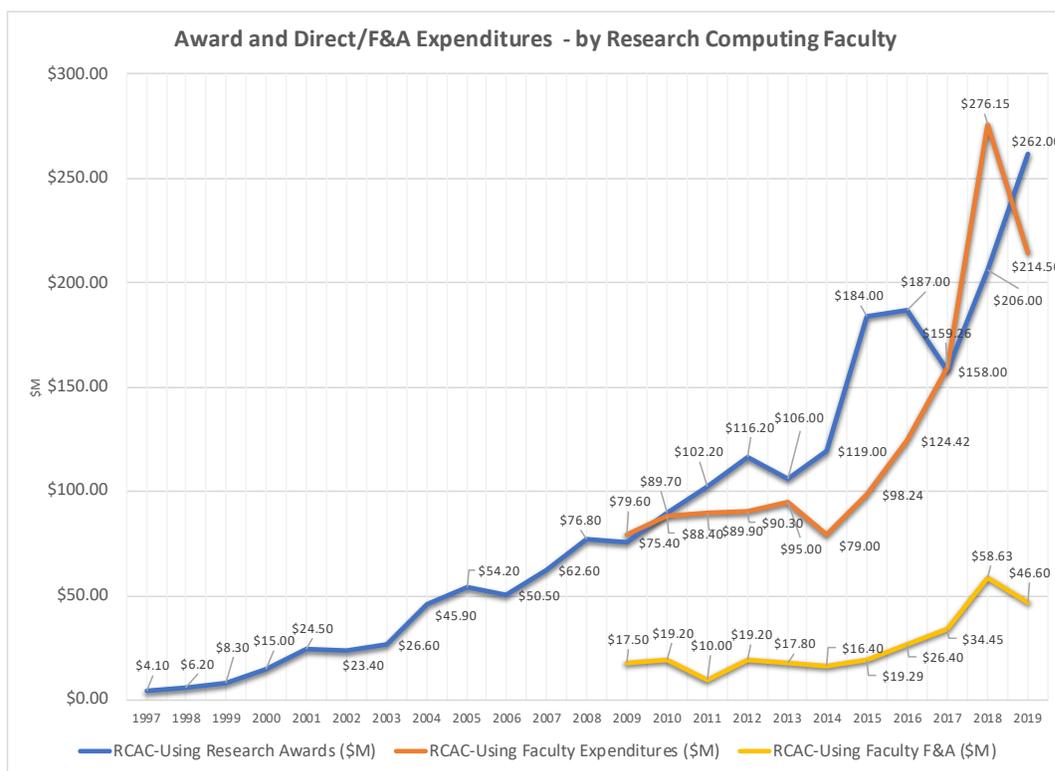
has grown to a program that is viewed as the reference model for campus computing nationwide, serving over **200 faculty partners**, and deploying 18 major systems since *Hamlet*.

By sharing the expenses of advanced computing between the university and faculty partners, the community cluster program has

delivered **high-performance computing at the highest proven value** to campus, enabling transformative research with systems like *Conte*, which was the 28th most powerful supercomputer in the world when it was deployed!

Now in 2019, research computing resources serve as a critical scientific instrument used by faculty from 59 departments from every academic college and all three Purdue campuses, and are increasingly used as one of Purdue's strengths with which to recruit and retain top faculty.

Research Computing supports **Purdue Moves**, enabling Engineering Preeminence at Scale, Computer Science, Computational Drug Discovery, Plant Sciences, and Life Sciences, and by supporting Data Science Education and Research. We're proud to say that we play a role



# DIRECTOR'S WELCOME

supporting over half of the \$520 million in sponsored research that was awarded to Purdue last year. Research computing is more than just clusters, however, and we support large-scale data storage services, provide advanced computational expertise to faculty, and operate a multi-million dollar sponsored research portfolio in cyberinfrastructure.

Cutting-edge discovery at Purdue requires constant evolution of the supporting infrastructure, and this past year has seen new deployments and expansions of multiple capabilities:

- The Gilbreth GPU cluster, supporting AI and machine learning applications.
- The Weber cluster, to support research requiring compliance with ITAR and NIST SP 800-171 regulations
- An expansion and refresh of the Research Data Depot
- Deployment of Box Cloud storage to support research, including HIPAA data.
- New capabilities on the Scholar cluster, which supports computational and data science education.
- Improved interactive computing capabilities on all resources

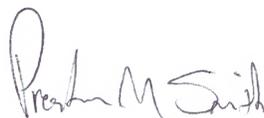
FY 2019 also saw success in sponsored research. Research computing PIs led or partnered in many funded projects of our own, earning nearly \$11M since 2017. Highlights include:

- NSF CSSI (GeoEDF): \$4.5M, 5 years
- NSF DIBBS (DEEDS): \$3.5M, 4 years
- NSF CC\* (Networking for big data instruments): \$323k, 2 years
- NSF CICI (Cybersecurity, network security): \$741k, 2 years
- NSF CICI (Framework for regulated research data): \$600k, 2 years
- NSF Cybercorps (Cybersecurity): \$350k, 3 years
- NSF Cybertraining (FAIR Science): \$500k, 3 years

In addition to research, the use of computing is becoming more and more critical to Purdue students, and we are at the forefront of efforts to teach **data science for all**.

Research Computing partners' use of the resources we make available covers all areas of Purdue. Our partners **model and simulate** everything from semiconductors, jet engines and weather and climate, to chemical structures and viruses. They **analyze** data from instruments, from microscopes on campus all the way to the Large Hadron Collider, as well as genetic data, real estate transactions, highway sensors, and plant imagery. They **visualize** augmented reality in classrooms, and much more.

I invite you to browse through this annual report, and learn in greater detail about all the discoveries enabled by Purdue's campus cyberinfrastructure.



Preston Smith  
Executive Director, Research Computing

**373**  
**MILLION**  
**COMPUTATIONAL**  
**HOURS**  
**DELIVERED IN**  
**2018**

# RESEARCH HIGHLIGHTS

## Purdue professor studies vortex dynamics using Brown supercomputer

Carlo Scalo, assistant professor of mechanical engineering, uses the Brown supercomputer for his work in aerodynamics and vortex dynamics. Scalo studies the flow of everything from a comparatively low-speed commercial airliner to a high-speed missile. The higher the speed, the more computational power is required to capture the details of the flow. Scalo, who also has access to even more powerful supercomputers at the Department of Defense Supercomputing Resource Center, finds himself needing to use them less often as Purdue continues to invest in on-campus computing resources.

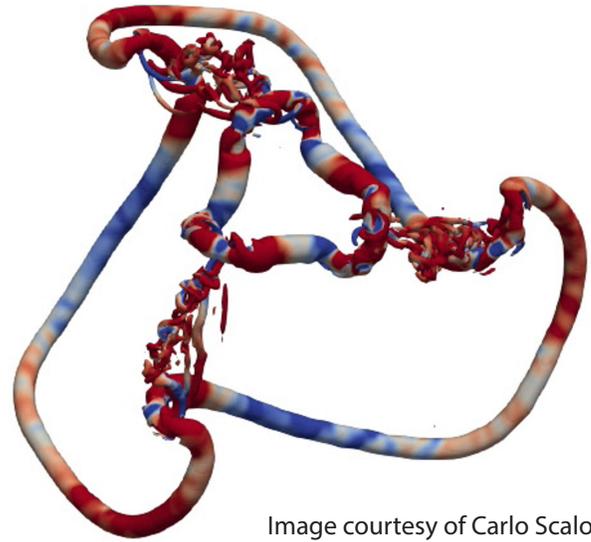


Image courtesy of Carlo Scalo

*“The scope of what on-campus computing is able to do for my research has really grown.”*

**Carlo Scalo, Purdue assistant professor of mechanical engineering**

Like Scalo, 38 percent of the faculty partners in Brown are at the rank of assistant professor. These early career faculty weigh the value of the community cluster program in their decision to come to Purdue.

Since 2012  
THE SHARE OF  
COMMUNITY CLUSTER  
PARTNERS AT THE  
RANK OF ASSISTANT  
PROFESSOR HAS  
GROWN FROM  
**5 to 38  
PERCENT**

# RESEARCH HIGHLIGHTS

## New Gilbreth all-GPU supercomputer has helped Purdue faculty make leaps in machine learning, artificial intelligence

Purdue faculty working in machine learning, artificial intelligence and other fields that are optimized for computations run on GPUs have a powerful resource in Gilbreth, Purdue's newest all-GPU supercomputer. Gilbreth's impact extends far beyond the 37 researchers who have purchased access to the cluster. It's playing an important role in the success of Purdue's Integrative Data Science Initiative (IDSI), a Purdue-wide initiative to help support and coordinate research, education and broader engagement in data science at Purdue, with a goal of bringing "data science to all" and ensuring that every Purdue graduate understands data science. Gilbreth has been used by the Data Science Consulting Service, a component of the IDSI that offers Purdue researchers hands-on support with data analysis.

*"Gilbreth is an invaluable resource for researchers on campus that makes a big difference to what we are able to accomplish."*

**Sunil Prabhakar, professor of computer science and Integrative Data Science Initiative director**

The supercomputer is named after Lillian Moller Gilbreth, an industrial engineer and efficiency expert, who became Purdue's first female engineering professor when she joined the faculty in 1935. Her research focused on combining psychology and engineering to improve efficiency in the workplace and home, and she pioneered the field now known as ergonomics.



**Lillian Gilbreth**

**196** FACULTY PARTNERS  
FROM **3** PURDUE CAMPUSES  
FROM ALL OF PURDUE'S PRIMARY COLLEGES  
**AND 59** DIFFERENT DEPARTMENTS  
HAVE USED THE COMMUNITY CLUSTER PROGRAM

# RESEARCH HIGHLIGHTS

## Recently retired Conte cluster drove research in nanotechnology, many other fields

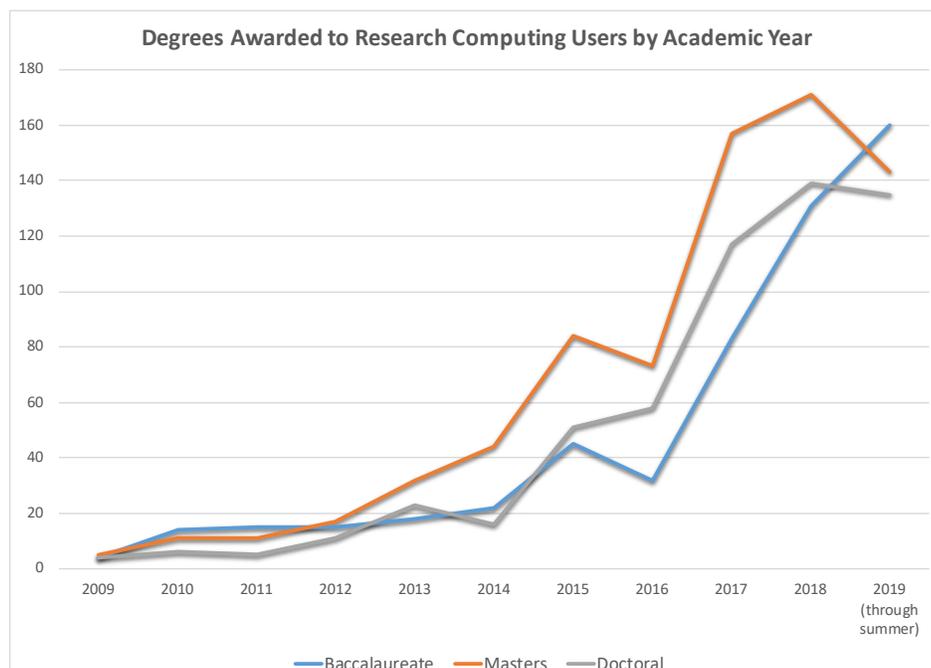
Purdue's powerful Conte supercomputer, which retired in 2018 after five years of service, was crucial to the development of powerful nanotechnology tools. When it was built in 2013, Conte replaced its predecessor Carter as the fastest supercomputer solely for use by researchers on a single campus and not part of a national supercomputing center. It debuted as number 28 on the Top500 list of the world's most powerful supercomputers, and remained on the list for its entire lifespan, clocking in at number 282 on the June 2018 ranking. Using Conte, a Purdue research group developed the Nanoelectronics Modeling (NEMO) suite, a set of simulation tools optimized for high-performance computing.



Conte also served as the computational backend for NanoHUB, a nanotechnology research platform led by Purdue's Network for Computational Nanotechnology. The supercomputer was named after Samuel Conte, who helped establish Purdue's first-in-the-nation computer science department.

***“Conte was simply essential for all this development.”***

**Tillmann Kubis, research assistant professor of electrical and computer engineering**



# RESEARCH HIGHLIGHTS

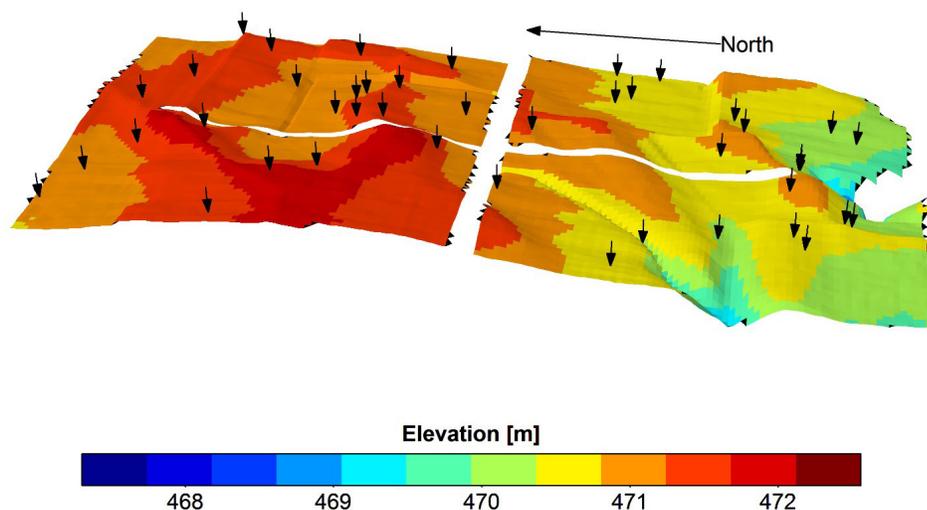
## Purdue soil scientist digs into data with the help of Data Workbench tool

Jason Ackerson, an assistant professor of agronomy, uses techniques such as spectroscopy to study the properties of different soils, and generates soil maps that can be fed into simulations such as climate models or hydrologic models. To clean up his data and estimate the accuracy of the resulting soil maps, he needs to run computationally intensive models. That's where Data Workbench – an interactive computing environment that provides access to web-based data analysis tools such as JupyterHub and R Studio Server – comes in. Instead of leaving his personal machine running all weekend, and hoping the computations have successfully finished by Monday morning, he can do them in a matter of a few hours on Data Workbench.

**IN 2018**  
**RESEARCH**  
**COMPUTING**  
**FACULTY PARTNERS**  
**PRODUCED OVER**  
**1,100**  
**PEER-REVIEWED**  
**PUBLICATIONS**

*“Data Workbench has taken something that was a real chore, and it’s now a trivial task. It’s made the science a lot faster.”*

**Jason Ackerson, assistant professor of agronomy**



# RESEARCH HIGHLIGHTS

## Envision Center celebrates 15th anniversary, unveils new virtual reality environment

Purdue's Envision Center, which was founded to help researchers visualize scientific data in a more intuitive and immersive way, celebrated its 15th anniversary in April 2019. At an open house marking the anniversary, the center unveiled The Forge, a new collaborative environment that allows multiple people to share the same virtual or augmented reality experience. The Envision Center's staff and student employees use the latest technology to collaborate with clients to create virtual reality and data visualization tools for research and educational use, collaborate on grant proposals and develop promotional media such as animated videos. The center's clients include 57 Purdue faculty members in 36 different academic departments.



# RESEARCH HIGHLIGHTS

## Weber cluster dedicated to high-performance computing with controlled research

Purdue's new Weber cluster gives faculty doing research subject to export control regulations, such as EAR or ITAR, or requiring compliance with NIST SP 800-171 standards, a powerful computational resource on campus. Weber will position Purdue to be an even better partner to industry and the federal government, and will allow the university to grow the amount of controlled research done here, says Mary Millsaps, Purdue's director of research information assurance.



*From left to right: Carolyn Ellis, Research Computing program manager for controlled research; Astronaut Mary Ellen Weber; Theresa Mayer, Purdue's executive vice president for research and partnerships; Preston Smith, Research Computing executive director.*

Weber is a key part of Purdue's Research Ecosystem for Encumbered Data, known as REED+, which will eventually include storage and computational tools for a variety of different kinds of regulated data.

Appropriately enough for a cluster that will enable cutting-edge national defense and space-related research, Weber is named for Astronaut Mary Ellen Weber, a Purdue alumna, engineer, scientist, and skydiver, who flew on two Space Shuttle missions, including one to the International Space Station. Weber visited campus over Homecoming for Purdue's astronaut reunion, where she toured the Research Computing data center.

## Purdue's Box research storage service a key component of Eli Lilly collaboration

Researchers participating in Purdue's five-year, \$52 million life sciences collaboration with Eli Lilly & Co. are relying exclusively on Purdue's instance of Box.com, which is compliant with regulations governing the storage of health data, including HIPAA and HITECH, to help securely store and share their regulated health data.

The Lilly collaboration was a driving force behind Purdue's agreement with Box, a cloud storage and collaboration service. The collaboration with Lilly began with a focus on improving delivery of injectable medicines and developing predictive models for clinical success to reduce the risk of investing in drug development.

**500**  
**RESEARCH**  
**LABORATORIES**  
**HAVE PURCHASED OVER**  
**2.5PB**  
**OF STORAGE SPACE IN**  
**PURDUE'S RESEARCH**  
**DATA DEPOT**

# EDUCATION AND OUTREACH

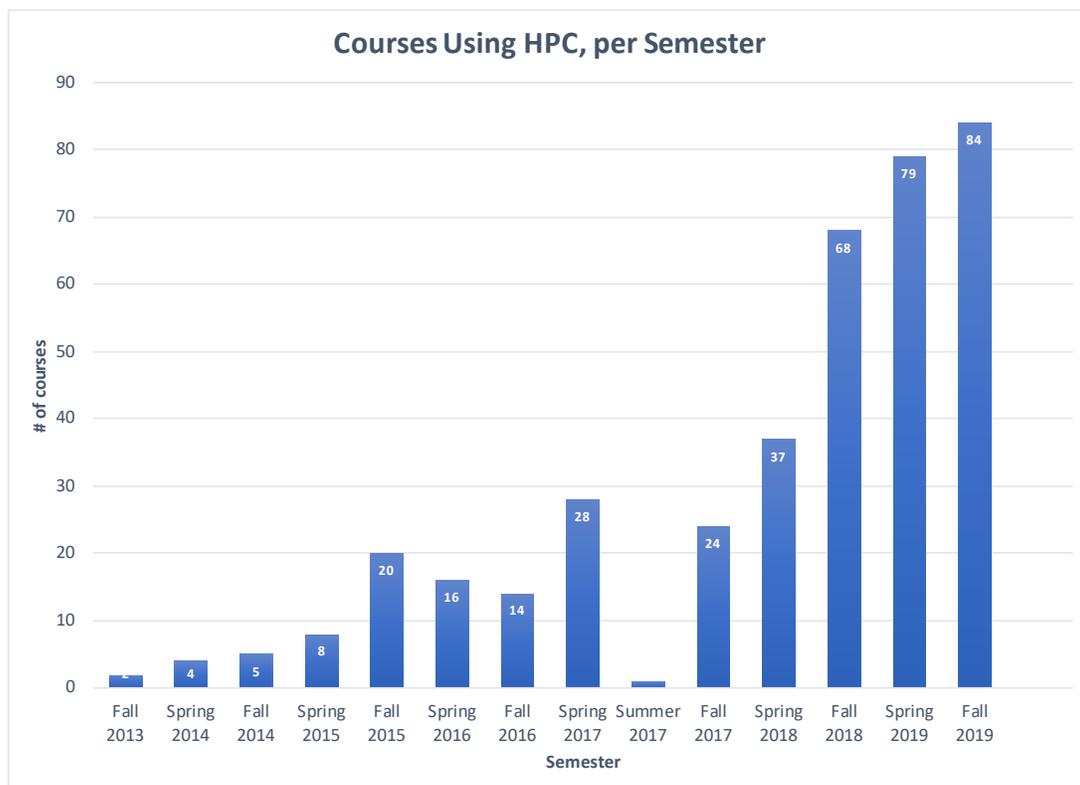
## Scholar cluster now features GPUs, increased storage

Purdue students using Scholar, a supercomputer Research Computing makes available for classroom use, can now use GPUs to study subjects such as machine learning, cryogenic electron microscopy and data science. Wen Jiang, a professor of biological sciences, is using the new GPU-enabled Scholar to teach undergraduate and graduate classes that involve cryo-EM image processing and 3-D reconstruction of the structures of macromolecular complexes and viruses. In the past, Jiang has taught his classes using Scholar's traditional CPUs, but as the image-processing software he uses increasingly becomes optimized for GPUs, having access to GPU nodes is crucial to address complex problems and increase the speed of computation.

In addition, Scholar now has increased storage capabilities, thanks to support from The Data Mine, a Purdue learning community that introduces students in a variety of different fields to concepts in data science.

***“The addition of GPUs to the Scholar cluster will allow my students to keep up with the fast-progressing cryo-EM field.”***

**Wen Jiang, professor of biological sciences**



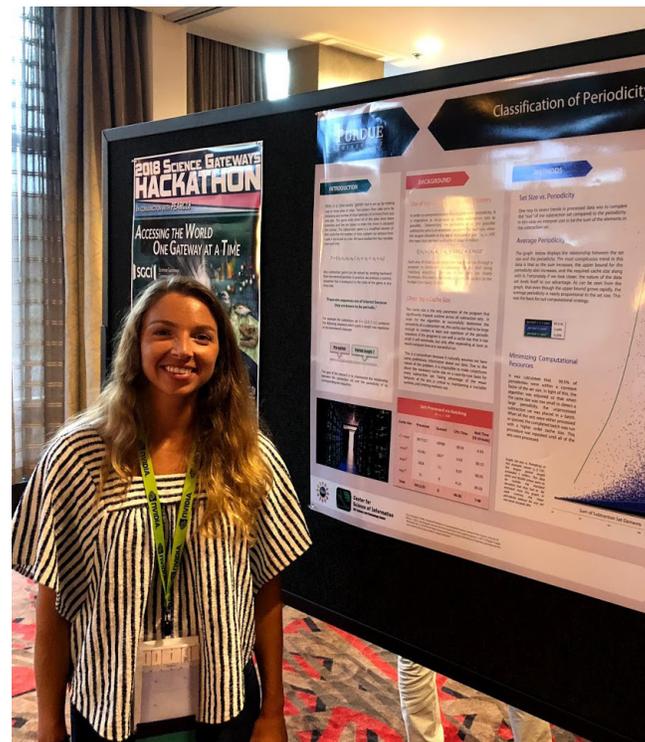
# EDUCATION AND OUTREACH

## Purdue Women in HPC member wins award at PEARC18

Nicole Brewer, then a senior in mathematics and computer science, and a member of Purdue's Women in HPC group, received the Phil Andrews Award at the 2018 Practice and Experience in Advanced Research Computing (PEARC) conference for her work on combinatorial game theory using Purdue's Halstead community cluster. She now works as a software engineer with Research Computing.

*“I walked away with more technical and soft skills than I could have ever imagined.”*

**Nicole Brewer, PEARC18 award winner**



## Purdue student-built intrusion detection system wins best student paper award at PEARC19

A team of students who deployed an intrusion detection system on Purdue's research network was awarded the Best Student Paper in the Facilitation of Advanced Research Computing track at the 2019 PEARC conference. The detection system, known as PULSAR, aligned Purdue's research cybersecurity framework with industry standards, making it easier for faculty to work with industries such as aerospace or defense. It also gave the students hands-on experience with research computing and cybersecurity.



# EDUCATION AND OUTREACH

## Purdue's first-ever all-women student supercomputing team competed at SC18

Young women interested in pursuing computer science have six new role models: the members of Purdue's first all-women student supercomputing team. The team, known as Ada Six in honor of Ada Lovelace, competed in the student cluster challenge at the SC18 supercomputing conference. They were advised by Betsy Hillery, Research Computing

manager for high-performance computing services, with support from other Research Computing staff members. With help from sponsors Dell, EMC, Nvidia and Mellanox, the students built their own supercomputer that they used to run a variety of real-world scientific applications during the 48 hour challenge. Being part of an all-women squad has been a special experience, the team members say.

***“It gives us a closer bond, because we have the shared experience of being women in science and technology.”***

**Heidi Anderson, junior in computer science from Saratoga, California**



The members of Ada Six, Purdue's first all-women student supercomputing team, with Donna Cumberland (left), former executive director of Research Computing, and U.S. Department of Energy Under Secretary of Science Paul M. Dabbar (second from left).

# EXTERNAL FUNDING

## Research Computing senior research scientist awarded \$4.6 million grant to build platform for geospatial data management

A team led by Research Computing senior research scientist Carol Song was awarded a five-year, \$4.6 million grant from the National Science Foundation to build a “plug and play” platform to allow researchers to easily access and process geospatial data. Song describes the platform, known as GeoEDF, as a successor to the Geospatial Data Analysis Building Blocks (GABBs), a project Song led that developed web-based geospatial data visualization, analysis and modeling tools and made them accessible to users on the science gateway MyGeoHub.

Despite the advent of geospatial data processing tools accessible even to non-programmers, data challenges remain in this area. Many geospatial data repositories lack standard interfaces and don’t provide data in a way that researchers can immediately use. Moreover, as field sensors become increasingly common, large volumes of streaming data are created, including so-called “crowdsourced” data generated by citizen scientists. GeoEDF’s data processing pipeline will help researchers retrieve and process only the data they need, and transform it into standardized formats.

**50%** OF PURDUE’S FISCAL YEAR 2019 GRANT AWARDS (\$262 MILLION) WERE AWARDED TO FACULTY USING RESEARCH COMPUTING RESOURCES

# EXTERNAL FUNDING

## Research Computing-led team awarded \$300,000 NSF grant to get big data from scientific instruments to computational, storage resources faster, more reliably

A project led by Research Computing staff was awarded a two-year, \$323,000 grant from the National Science Foundation to build a high-speed network infrastructure that will quickly and reliably get big data from scientific instruments to storage and computational resources. Carol Song, senior research scientist, is the project's principal investigator and Preston Smith, research computing executive director, is a co-PI.

Researchers using scientific instruments increasingly generate very large amounts of data, and moving that data to places such as the Data Depot storage array or Purdue's community cluster supercomputers for storage or computation requires tapping into high-speed research networks. To make that easier, this project is upgrading certain campus facilities with greater bandwidth, giving researchers the ability to connect at much faster speeds. The upgrades will not only provide a faster network connection, but will use the "Science DMZ" model to provide a more predictable connection that is optimized for scientific data flows.

**52%**  
**OF PURDUE'S**  
**FISCAL YEAR 2018**  
**RESEARCH EXPENDITURES**  
**(\$214 MILLION)**  
**WERE EXPENDED BY**  
**FACULTY PARTNERS**

***"We're getting data from your instrument to where you need to work on it quicker and more reliably, and decreasing your time to science."***

**Preston Smith,**  
**Research Computing executive director**

# EXTERNAL FUNDING

## Purdue team awarded NSF grant to generate framework for working with regulated data

Carolyn Ellis, Research Computing program manager, is a co-principal investigator on a two-year, almost \$600,000 grant from the National Science Foundation to develop a regulated data research framework that builds on and extends Purdue's resources. The new system, which will be known as the Research Ecosystem for Encumbered Data (REED+), will bring together staff from across campus to develop best practices for working with regulated data and to make sure that compliance with regulated data requirements is achieved through every part of the research lifecycle.

The team will use undergraduate cybersecurity students to accomplish much of the work. The students will work with Research Computing staff to develop training and educational materials, as well as an awareness campaign about the handling of regulated data. The REED+ framework will eventually be shared with the national IT and research administration communities and is likely to serve as a model for other universities.

***“This award was a real indication that Purdue has become nationally recognized for research data security.”***

**Mary Millsaps, Purdue's director of research information assurance**

**RESEARCH COMPUTING STAFF  
WERE PIS OR CO-PIS ON  
7 GRANTS  
THAT BROUGHT IN MORE  
THAN \$7.2 MILLION IN  
FUNDING IN FY2019**

# CONTACTS

## **Preston Smith**

Executive Director, Research Computing  
psmith@purdue.edu  
765-494-9729

## **Carol Song**

Director of Scientific Solutions  
carolxsong@purdue.edu  
765-496-7467

## **George Takahashi**

Envision Center Technical Director  
gtakahas@purdue.edu  
765-496-7888

[www.rcac.purdue.edu](http://www.rcac.purdue.edu)

Image courtesy of Hector Gomez, Purdue associate professor of mechanical engineering, who uses Purdue's supercomputers for his work in the modeling and simulation of biomechanics problems, including tumor growth.

**PURDUE**  
UNIVERSITY

Information Technology

Purdue University is an EA/EOU