SAN DIEGO SUPERCOMPUTER CENTER at UC SAN DIEGO

GATEWAY to Data-Driven Discovery

Providing an End-to-End Data Computing Ecosystem

SDSC

ANNUAL REPORT FY2017/18

SAN DIEGO SUPERCOMPUTER CENTER

As an Organized Research Unit of UC San Diego, SDSC is considered a leader in data-intensive computing and cyberinfrastructure, providing resources, services, and expertise to the national research community, including industry and academia. Cyberinfrastructure refers to an accessible, integrated network of computer-based resources and expertise, focused on accelerating scientific inquiry and discovery. SDSC supports hundreds of multidisciplinary programs spanning a wide variety of domains, from earth sciences and biology to astrophysics, bioinformatics, and health IT. SDSC's petascale *Comet* supercomputer continues to be a key resource within the National Science Foundation's (NSF) eXtreme Science and Engineering Discovery Environment (XSEDE) program.

SDSC INFORMATION

Michael L. Norman, Director

San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive MC 0505
La Jolla, CA 92093-0505
Phone: 858-534-5000

info@sdsc.edu
www.sdsc.edu

Jan Zverina
Division Director, External Relations
jzverina@sdsc.edu
858-534-5111

GATEWAY TO DISCOVERY

SDSC Annual Report FY2017/18
(PDF version available online at the SDSC website)

EDITOR:
Jan Zverina

CO-EDITOR:
Warren Froelich

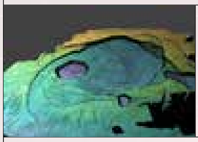
CONTRIBUTORS:
Warren Froelich, Julie Gallardo, Ron Hawkins, Fritz Leader, Susan Rathbun, Bob Sinkovits, Shawn Strande, Ben Tolo, Nancy Wilkins-Diehr, Nicole Wolter, Jan Zverina

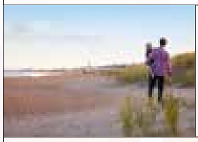
CREATIVE DIRECTOR:
Ben Tolo

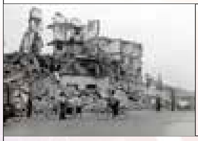
PHOTOGRAPHY:
Jon Chi Lou

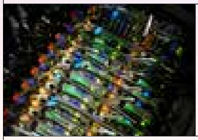
All financial information is for the fiscal year ended June 30, 2018. Any opinions, conclusions, or recommendations in this publication are those of the author(s) and do not necessarily reflect the views of NSF, other funding organizations, SDSC, or UC San Diego. All brand names and product names are trademarks or registered trademarks of their respective holders.

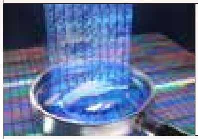
© 2018 The Regents of the University of California











DIRECTOR'S LETTER

Delivering a Lasting IMPACT

> *In an era now transformed by the sheer volume of scientific data... we are now trying to sift through billions, trillions, and even quadrillions of bits and bytes of electro-magnetic code.*

During the first quarter of 2018 SDSC joined UC San Diego Chancellor Pradeep Kohsla in welcoming the Halicioğlu Data Science Institute (HDSI) and the news that SDSC will be one of the homes for this new campus-wide initiative. All of us welcome the Institute and the HDSI faculty, and look forward to leveraging SDSC's expertise and resources in high-performance computing and data-enabled science as we integrate the needs of campus into our strategic initiatives.

As a top research university, UC San Diego is leading the way in establishing data-enabled science as a key part of its curriculum, and I view SDSC's reputation for computing-at-scale as a key support in achieving this goal. SDSC is well-positioned to collaborate with HDSI to develop education and training opportunities for UC San Diego students and postdocs. We also expect that our *Comet* supercomputer – which recently received a supplemental National Science Foundation (NSF) award for a sixth year of service into 2021 – as well as many of our data-centric analytics platforms will be called upon to advance HDSI-based research in areas such as machine learning and artificial intelligence.

In an era now transformed by the sheer volume of scientific data being collected by large-scale scientific instruments around the world – interferometers, cyclotrons, sequencers, solenoids, and cryo-electron microscopes – we are now trying to sift through billions, trillions, and even quadrillions of bits and bytes of electro-magnetic code. Scientists believe that within these troves of information lie clues to answering questions that humankind has been asking for years, such as how the universe was formed, what is the underlying nature of matter, and what are the true causes of disease.

SDSC's Vision Statement
To deliver lasting impact across the greater scientific community by creating end-to-end computational and data solutions to meet the biggest research challenges of our time.

For this reason, the ability to convert all this digital data into meaningful discoveries has taken on added significance on both a national level as well as across UC San Diego. Indeed, one of the NSF's top "Big Ideas" for the future includes "Harnessing the Data Revolution."

All this aligns well with SDSC's own national mission of being at the forefront of creating an advanced cyberinfrastructure. Among other things, much of science requires the integration of computational resources in an "ecosystem" that includes sophisticated workflow tools to orchestrate complex pathways for scheduling, data transfer, and processing. Massive sets of data collected through these efforts also require tools and techniques for filtering and processing, plus analytical techniques to extract key information. Moreover, the system needs to be effectively automated across different types of resources, including instruments and data archives.

> *Armed with ever-more powerful large-scale scientific instruments, research teams around the globe… are already converging to build an impressive portfolio of scientific advances and discoveries, with supercomputers serving as the critical linchpin.*

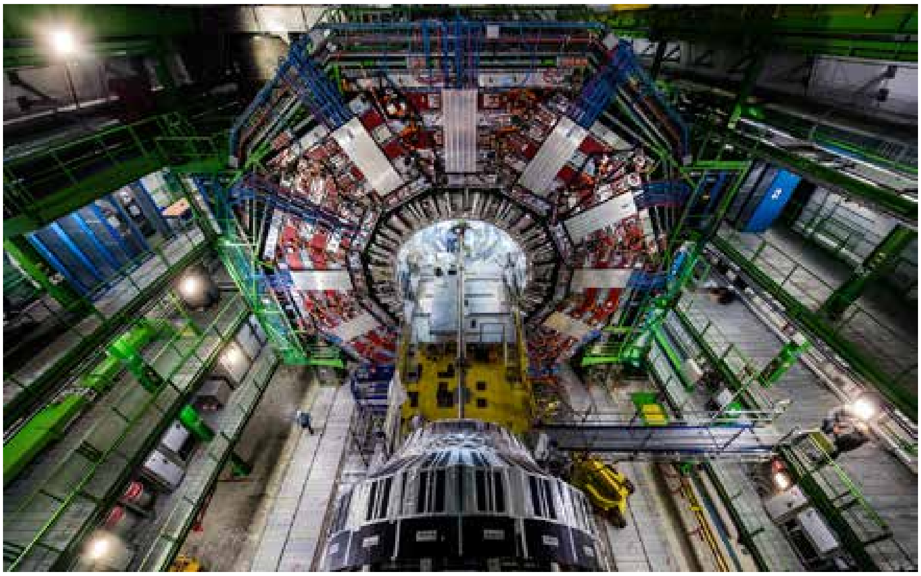
Armed with ever-more powerful large-scale scientific instruments, research teams around the globe – some encompassing a wide variety of disciplines – are already converging to build an impressive portfolio of scientific advances and discoveries, with supercomputers serving as the critical linchpin for all these investigations.

Moreover, advancing scientific discovery has always been about education and innovation, and recent events mean one thing: opportunities for all of us. By this, I mean opportunities to be collaborative on both the campus and UC fronts, as well as through our mission of helping to develop advanced cyberinfrastructure nationally to scale. As the scale of data and complexity of these experimental projects increase, it's more important than ever that centers such as SDSC respond by providing advanced HPC systems and expertise that become part of the integrated ecosystem of research and discovery.

One final thought: as data-driven research now underpins just about every area of scientific advancement, we are mindful of the need to focus on making sure the myriad software projects at SDSC that serve both science and society – Sherlock, Firemap, and SeedMe, just to name a few – are sustainable in the long-run not only through funding grants but through business models that include longer-term industry partnerships and other support mechanisms.

I now invite you to browse through our latest Annual Report to get a sense of the depth and breadth of research activity going on at SDSC.

Michael L. Norman
SDSC Director



Under a partnership between a team of UC San Diego physicists and the Open Science Grid (OSG), a multi-disciplinary research partnership funded by the U.S. Department of Energy and the NSF, SDSC's *Gordon* supercomputer provided auxiliary computing capacity by processing massive data sets generated by the Compact Muon Solenoid, one of two large general-purpose particle detectors at the Large Hadron Collider used by researchers to find the elusive Higgs particle.

Image credit: CERN.

SDSC's 2018 'PI' PERSON OF THE YEAR

MEET KC CLAFFY

KC Claffy was unanimously chosen by an internal team as SDSC's 2018's π Person of the Year. Now in its fifth year, this award recognizes SDSC researchers who have one 'leg' in a science domain and the other in cyberinfrastructure technology, similar to the π symbol.

Claffy is founder and director of the Center for Applied Internet Data Analysis (CAIDA), based at SDSC. She leads CAIDA's research and infrastructure efforts in internet cartography, aimed at characterizing the changing nature of the internet's topology, routing and traffic dynamics, and investigating the implications of these changes on network science, architecture, infrastructure security and stability, and public policy.

Claffy was named the 2017 recipient of the prestigious Jonathan B. Postel Service Award. The Internet Society – a global non-profit dedicated to ensuring the open development, evolution and use of the internet – called Claffy a "visionary in the field of internet research." An adjunct professor of computer science and engineering at UC San Diego, Claffy has been at SDSC since 1991 and holds a Ph.D. in Computer Science from the university.

"Claffy is a true innovator on either side of the 'Pi' equation," said SDSC Director Michael Norman. "She is a leading expert in internet measurement, while relying on data management and analytics cyberinfrastructure now necessary for such data-intensive research."

Today's internet, which began life in the late 1960s as a government-funded experiment never intended for commercial use, is composed of about half-million independent units of routability and some 60,000 independently autonomous networks. Claffy says the biggest issue facing the internet today is security, and its future lies in accurate measurement through multiple sources of data. The issue of 'Net Neutrality' is also on the horizon.

Net neutrality has become a little bit of a political football, says Claffy, adding that we don't even have close to enough data to truly litigate the net neutrality debate in an equitable way. “We don’t know enough. You need measurement.”

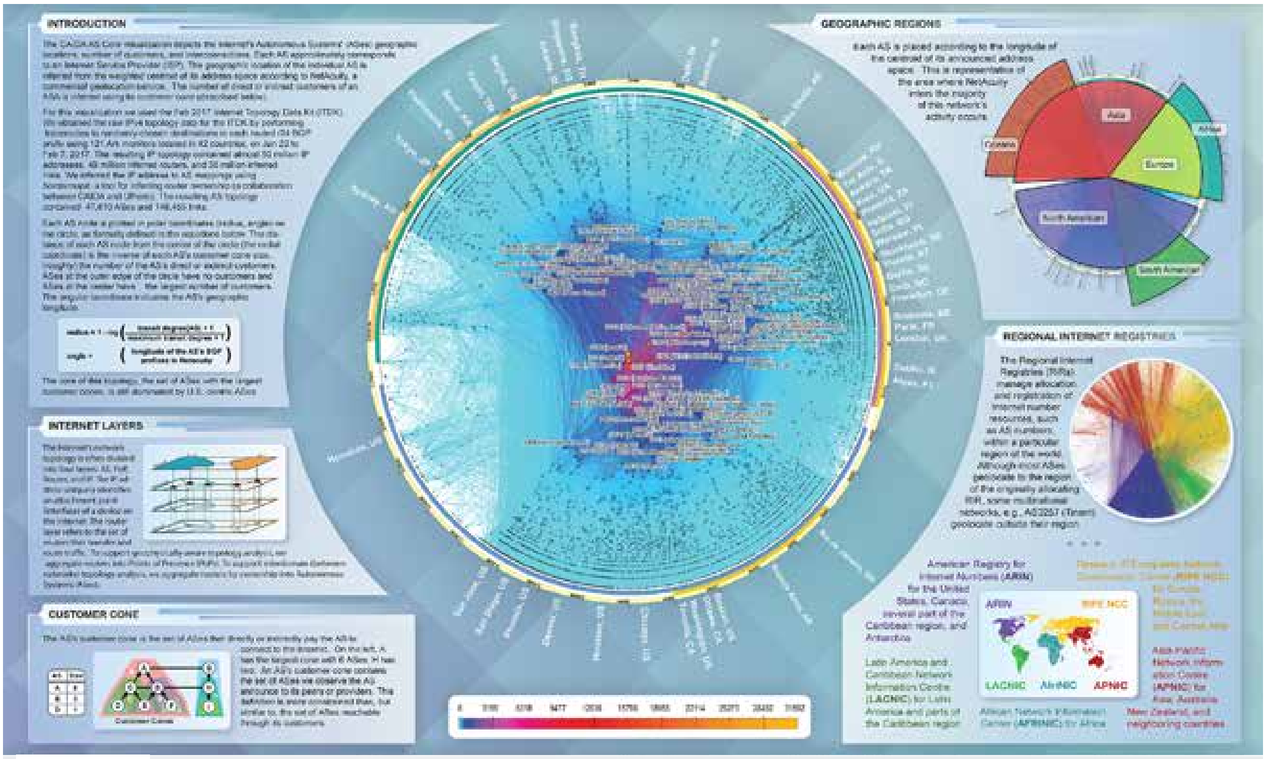
Toward that end, CAIDA was recently awarded a $4 million, five-year National Science Foundation (NSF) grant to integrate several of its existing research infrastructure measurement and analysis components into a new Platform for Applied Network Data Analysis, or PANDA. The platform, to include a science gateway component, is in response to feedback from the research community that current modes of collaboration do not scale to the burgeoning interest in the scientific study of the internet.

“Eventually I think society will need something like a bureau of internet statistics, some neutral, sustainable entity for data-gathering, curation, analysis, presentation, and sharing to support reproducible scientific studies, informed policy, and even compliance efforts,” says Claffy. “What we hope to prototype is a platform that tackles both technical and privacy challenges to demonstrate that a diverse set of users can answer important, complex questions about the internet using data that can be gathered and integrated from various sources.”

Claffy notes that it helps to be at UC San Diego, a campus with a rich tradition of interdisciplinary research, and where SDSC offers high-performance computing facilities and expertise to the CAIDA research group. “And of course we’re honored to have received over 20 years of federal R&D funding — grant by grant and contract by contract — most significantly from the NSF and the Department of Homeland Security, to support the research and infrastructure components that will enable this next step in CAIDA’s history.” Read more about CAIDA on page 27.

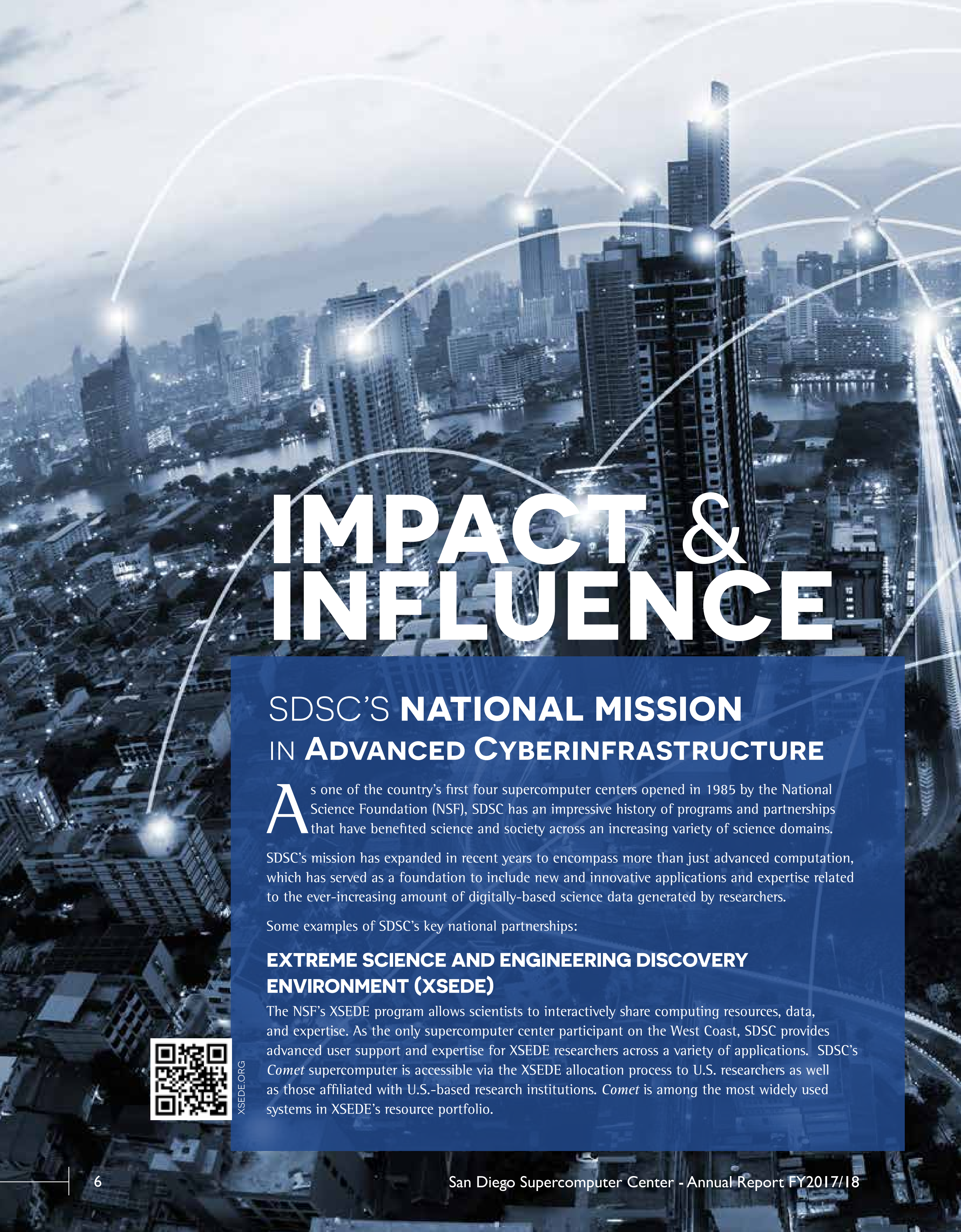
PANDA
GOO.GL/YPCQCE

MEASURING THE INTERNET
YOUTU.BE/A1RNA4LA1ZE



Since 2000, CAIDA has generated internet topology maps in order to visualize the shifting topology of the internet over time. These maps depict the internet’s autonomous systems’ geographic locations, number of customers, and interconnections. View the full-size current and past versions online using the QR/URL to the left. *Image courtesy of CAIDA.*

WWW.CAIDA.ORG/RESEARCH/
TOPOLOGY/AS_CORE_NETWORK/



IMPACT & INFLUENCE

SDSC'S NATIONAL MISSION IN ADVANCED CYBERINFRASTRUCTURE

As one of the country's first four supercomputer centers opened in 1985 by the National Science Foundation (NSF), SDSC has an impressive history of programs and partnerships that have benefited science and society across an increasing variety of science domains.

SDSC's mission has expanded in recent years to encompass more than just advanced computation, which has served as a foundation to include new and innovative applications and expertise related to the ever-increasing amount of digitally-based science data generated by researchers.

Some examples of SDSC's key national partnerships:

EXTREME SCIENCE AND ENGINEERING DISCOVERY ENVIRONMENT (XSEDE)

The NSF's XSEDE program allows scientists to interactively share computing resources, data, and expertise. As the only supercomputer center participant on the West Coast, SDSC provides advanced user support and expertise for XSEDE researchers across a variety of applications. SDSC's *Comet* supercomputer is accessible via the XSEDE allocation process to U.S. researchers as well as those affiliated with U.S.-based research institutions. *Comet* is among the most widely used systems in XSEDE's resource portfolio.

XSEDE.ORG

OPEN SCIENCE GRID CONSORTIUM

OPENSCIENCEGRID.ORG

Open Science Grid Consortium (OSG) is a multi-disciplinary research partnership specializing in high-throughput computational services funded by the U.S. Department of Energy and NSF. Through a partnership with XSEDE, OSG scientists have access to resources such as *Comet* to further their research. The integration of *Comet* into the OSG provisioning system was led by a team including Frank Würthwein, an expert in experimental particle physics and advanced computation, and SDSC's lead for distributed high-throughput computing. Würthwein also serves as OSG's executive director. OSG operates services that allow for transparent computation across more than 150 computing clusters worldwide, including the use of National Grid Initiatives in Europe, Asia, and the Americas.

SUPPORTING THE NATIONAL BRAIN INITIATIVE THROUGH THE NEUROSCIENCE GATEWAY

WWW.NSGPORTAL.ORG

Charting brain functions in unprecedented detail could lead to new prevention strategies and therapies for disorders such as Alzheimer's disease, schizophrenia, autism, epilepsy, traumatic brain injury, and more. The BRAIN Initiative (Brain Research through Advancing Innovative Neurotechnologies), launched by President Barack Obama in 2013, is intended to advance the tools and technologies needed to map and decipher brain activity, including advanced computational resources and expertise.

NSF WEST BIG DATA INNOVATION HUB (WBDIH)

WESTBIGDATAHUB.ORG

The NSF supports four regional Big Data Innovation Hubs throughout the U.S. The Western region, comprised of 13 states with Montana, Colorado, and New Mexico marking the eastern boundary, is led from SDSC, UC Berkeley, and the University of Washington's eScience Institute. The Hub's purpose is to facilitate multi-state, multi-sector partnerships in the area of 'big data' innovation. The WBDIH regularly hosts workshops including Data Storytelling, Data Hackathon Best Practices, and the National Transportation Data Challenge. This past year, the WBDIH hosted and participated in several water-themed workshops, including the California Water Data Challenge. The WBDIH also launched its carpentry initiative, to expand the number of software and data carpentry trainers and raise awareness of this free training model to new regional areas. Part of this initiative included a data awareness event at Pala, a San Diego tribal community, in partnership with the Southern California Tribal Digital Village (SCTDV).

OPEN STORAGE NETWORK (OSN)

HTTPS://GOO.GL/KSMYRx

SDSC was one of four partners awarded a $1.8 million grant from the NSF in mid-2018 for a data storage network demonstration project. During the next two years the team will combine its expertise to develop the Open Storage Network (OSN), which will allow academic researchers across the nation to work with and share their data more efficiently than before. The project, led by Alex Szalay of Johns Hopkins University, is from Schmidt Futures, a philanthropic initiative founded by former Google Chairman Eric Schmidt. Christine Kirkpatrick is leading SDSC's efforts. OSN will benefit from integrating with previous NSF investments, such as the CC* project that brought 10/100Gbps connectivity to several U.S. universities. OSN leverages key data storage partners throughout the U.S. including the National Data Service (see below) and the four NSF-funded Big Data Regional Innovation Hubs.

NATIONAL DATA SERVICE (NDS)

NATIONALDATASERVICE.ORG

SDSC has taken a leadership role in NDS through SDSC's Christine Kirkpatrick, the organization's first executive director. Founded by a consortium of U.S.-based research computing centers, governmental agencies, libraries, publishers, and universities, NDS builds on the data archiving and sharing efforts already underway within scientific communities and links them together with a common set of services. NDS is a vision for how scientists and researchers across all disciplines can find, reuse, and publish data, while providing a workbench for research projects and a training platform for future curators, data stewards, and scientists – its value is in making it easier to use resources together through dependency management and increased interoperability between cyberinfrastructure services.

PROVIDING **SCIENCE GATEWAYS** FOR RESEARCHERS

Nancy Wilkins-Diehr is an associate director of SDSC and has served as co-PI of XSEDE (eXtreme Science and Engineering Discovery Environment) as well as director of XSEDE's Extended Collaborative Support Services. Her XSEDE responsibilities include providing user support for Science Gateways as well as education, outreach, and training.

In mid-2016, a collaborative team led by SDSC Associate Director Nancy Wilkins-Diehr was awarded a five-year, $15 million NSF grant to establish a Science Gateways Community Institute (SGCI) to accelerate the development and application of highly functional, sustainable science gateways that address the needs of researchers across the full spectrum of NSF directorates and other federal agencies. Science gateways make it possible to run the available applications on supercomputers such as *Comet* so results come quickly, even with large data sets. Moreover, browser access offered by gateways allows researchers to focus on their scientific problem without having to learn the details of how supercomputers work. SGCI provides the needed, cost-effective focal point for developers to share experiences, find expertise, and make available community-contributed tools. In April 2018, SGCI was approved by an NSF review panel to move into the execution phase.

SGCI external-facing activities during early 2018 included a workshop held in conjunction with the Molecular Sciences Software Institute at the American Chemical Society's national meeting and expo, as well as implementation of an intense, in-person boot camp program, held twice a year for five-day periods to help attendees form lasting bonds among peers who share common research challenges within similar science domains. In June 2018, two such boot camps were held, including one for the EarthCube all-hands meeting in Washington DC, and by request at the International Workshop on Science Gateways in Edinburgh, Scotland. "The response to these boot camps significantly surpassed the expectations of the entire SGCI team," said Wilkins-Diehr.

In the area of workforce development, SGCI is now preparing additional programs including a four-week coding institute, focused internships, and an SGCI hackathon. "Our core focus is connecting people and resources to accelerate discovery by empowering the entire science gateway community," said Wilkins-Diehr. "Our target market is U.S.-based academic and non-profit students, researchers, and educators who are eager to support their communities. While we make and measure progress one calendar quarter at a time by focusing on prioritized deliverables, SGCI's ultimate goal is to be an autonomous world-class leader and think tank for science gateways. With that we hope SGCI becomes a critical path to scientific discovery via a target of 100,000 publications and training future gateway talent by educating some one million students."

The Gateways conference series is hosted each fall at a location in the US. It offers gateway developers and users an opportunity to connect with and learn from colleagues with diverse experiences, approaches, and academic interests.

SCIENCEGATEWAYS.ORG

Science Gateways
Pioneered by SDSC Researchers

SDSC Biology Researcher Mark Miller is PI for the popular CIPRES Science Gateway.

CIPRES

One of the most popular science gateways across the entire XSEDE resource portfolio is the CIPRES science gateway, created as a portal under the NSF-funded CyberInfrastructure for Phylogenetic RESearch (CIPRES) project in late 2009. The gateway allows scientists to explore evolutionary relationships by comparing DNA sequence information between species using supercomputers provided by the NSF's XSEDE project.

In mid-2018, CIPRES was awarded more than $2.8 million in grants from the NSF and National Institutes of Health (NIH) that combined will expand its software and resource capabilities for biological research. Currently, CIPRES supports more than 10,000 researchers who are investigating a wide range of biological fields, including the evolutionary history of proteins, viruses, bacteria, plants, and animals; identifying new genera and species; evaluating and developing new techniques; and exploring the evolutionary history of diverse populations.

At least 4,500 peer-reviewed publications in journals have relied on CIPRES resources for their research, with results appearing in some of biology's most prestigious journals such as *Science*, *Nature*, *Proceedings of the National Academy of Sciences* (PNAS), and *Cell.*

"Understanding the evolutionary history of living organisms is a central goal of nearly every discipline in biology," said Mark Miller, the gateway's principal investigator based at SDSC. "We're gratified by CIPRES' wide acceptance by the biological community and pleased that both federal agencies have seen fit to support its future operations." Please read more about CIPRES-enabled research on page 40.

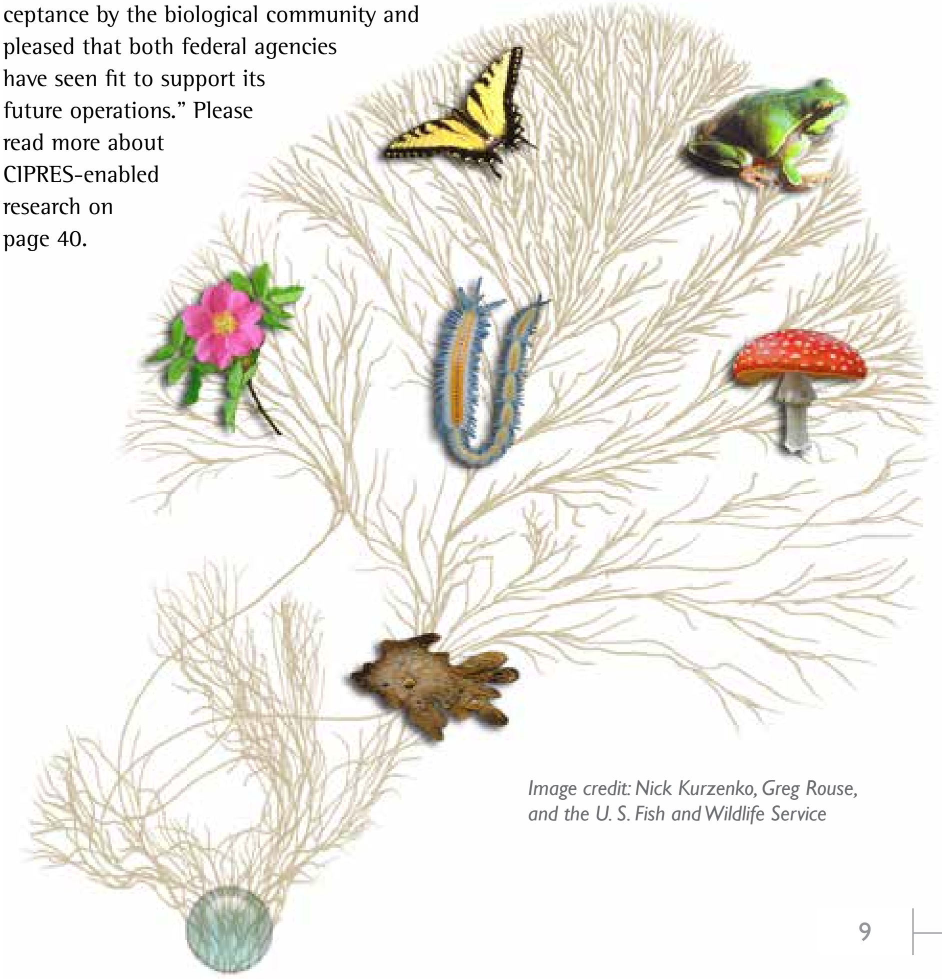


Image credit: Nick Kurzenko, Greg Rouse, and the U. S. Fish and Wildlife Service

WWW.PHYLO.ORG

NATIONAL IMPACT & INFLUENCE
SCIENCE GATEWAYS PIONEERED BY SDSC RESEARCHERS

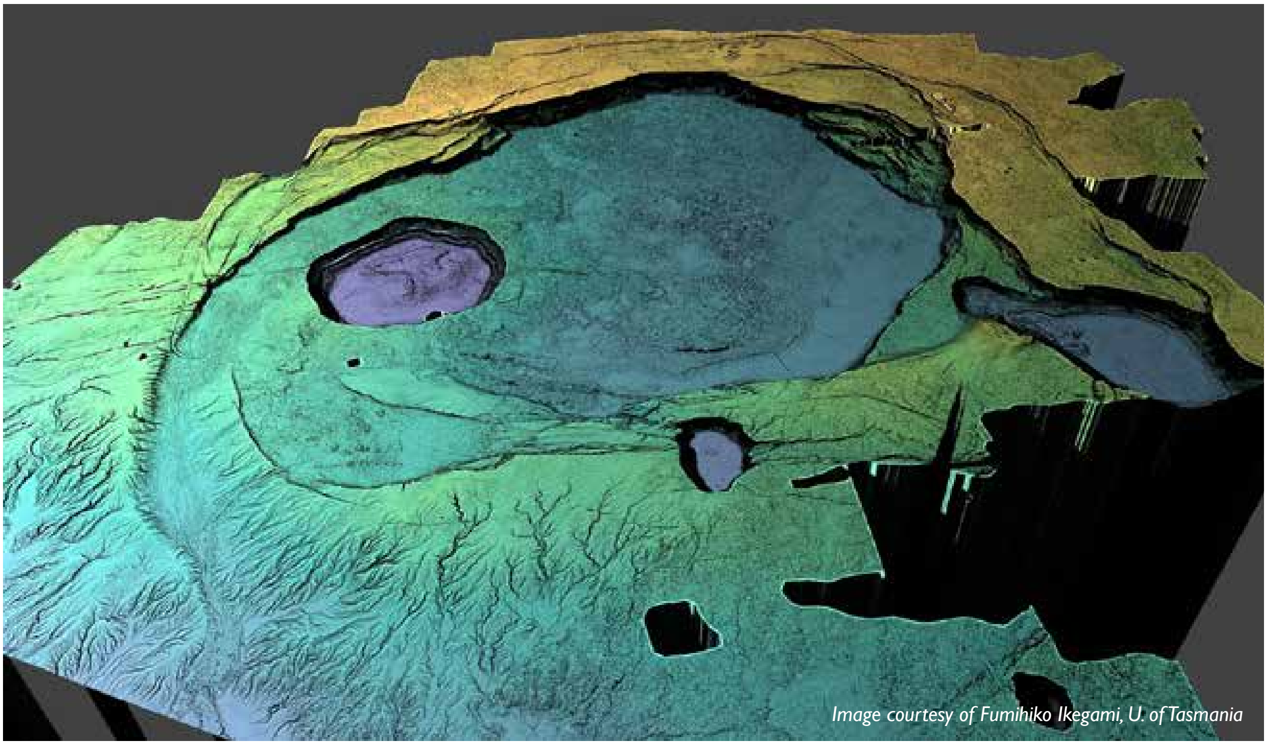


Image courtesy of Fumihiko Ikegami, U. of Tasmania

The Halema'uma'u crater at the center of the Kilauea volcano summit caldera, on the Big Island of Hawaii, was occupied by a lava lake from 2008 until the recent 2018 eruption episode. The shape of the crater has been changing over time through gravitational collapses and overflow events. The lava lake drained in 2018 in association with a series of explosive eruptions. Airborne lidar data, such as this 2009 dataset, provides a high-resolution snapshot of actively changing landforms and helps us to understand past and ongoing processes of the Earth.

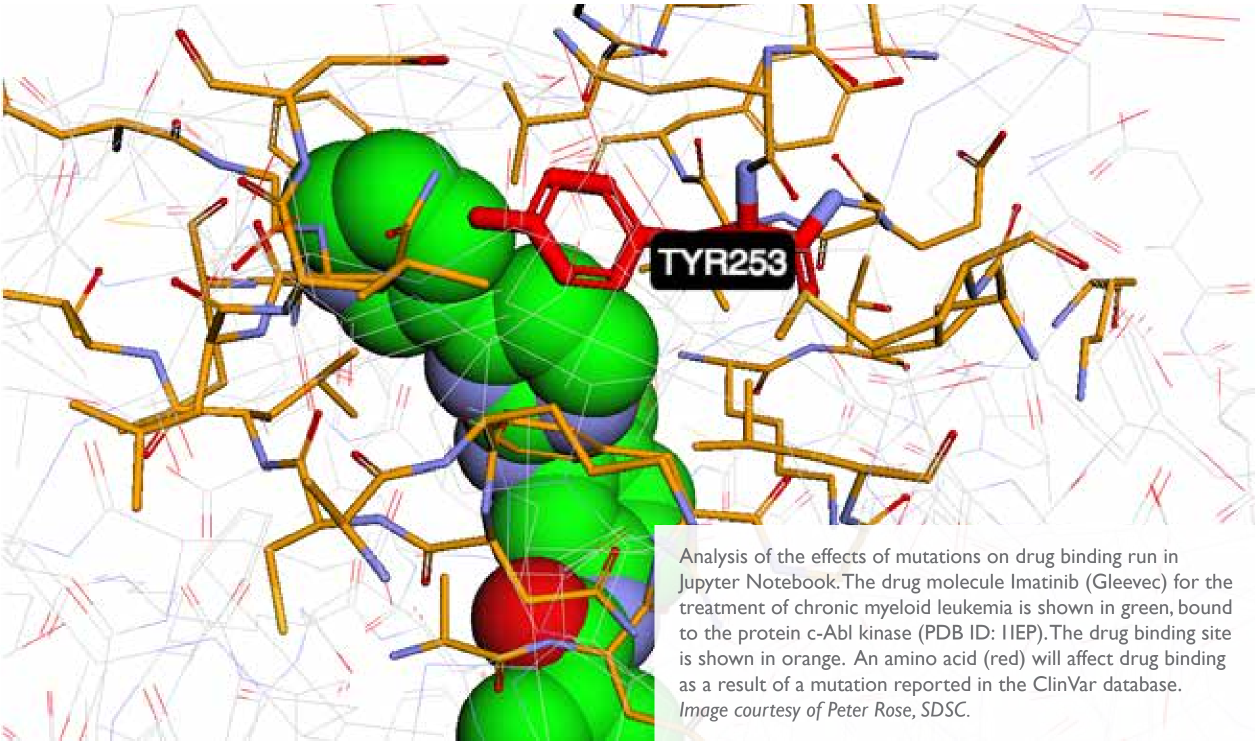
Viswanath Nandigam is PI and chief software architect for the Open Topography Facility. He is also associate director of SDSC's Advanced Cyberinfrastructure Development group (ACID).

OPENTOPOGRAPHY

Initiated in 2009 with funding from the NSF, OpenTopography provides easy access to earth science-oriented, high-resolution topographical data and processing tools for a broad spectrum of research communities. A collaboration between UC San Diego, Arizona State University, and UNAVCO, OpenTopography employs sophisticated cyberinfrastructure that includes large-scale data management, distributed high-performance computing (HPC), and service-oriented architectures, providing researchers with efficient web-based access to large, high-resolution topographic datasets.

Currently, OpenTopography data holdings comprise 277 lidar point cloud datasets with over 1.11 trillion lidar returns and 152 high-resolution raster datasets covering 219,648 km^2, and six global datasets including the highly popular Satellite Radar Topography Mission (SRTM) global 30m dataset. OpenTopography has established itself as resource that supports a broad interdisciplinary user base in academia and beyond. Since its initiation, over 75,450 unique users have run jobs via the gateway portal and it continues to see significant growth in usage, averaging close to 400 new user registrations per month. At least 290 peer-reviewed articles and other publications have been produced using OpenTopography resources. These include academic works in earth science, ecology, hydrology, geospatial and computer science, and engineering.

OPENTOPOGRAPHY.ORG



Analysis of the effects of mutations on drug binding run in Jupyter Notebook. The drug molecule Imatinib (Gleevec) for the treatment of chronic myeloid leukemia is shown in green, bound to the protein c-Abl kinase (PDB ID: 1IEP). The drug binding site is shown in orange. An amino acid (red) will affect drug binding as a result of a mutation reported in the ClinVar database. *Image courtesy of Peter Rose, SDSC.*

PROTEIN DATA BANK BENEFITS GLOBAL HEALTH, SCIENCE, ECONOMY

By mid-2018, the Protein Data Bank (PDB), the single worldwide repository for three-dimensional structures of large molecules and nucleic acids, archived more than 140,000 structures critical to research and education. Co-located at SDSC in conjunction with UC San Diego's Skaggs School of Pharmacy and Pharmaceutical Sciences; and Rutgers, The State University of New Jersey, the Research Collaboratory for Structural Bioinformatics (RCSB) PDB serves as the U.S. data center for the archive. PDB structures are carefully curated to bring the most accurate data possible into the PDB archive. The website at RCSB.org supports more than one million users representing a broad range of skills and interests. A variety of search and analysis tools provide access to structure data, comparative data, and external annotations such as information about point mutations and genetic variations. Fast, interactive 3D displays of molecular complexes containing millions of atoms – without plug-ins – is possible on desktop computers and even smartphones using the NGL (New Graphic Library) Viewer. NGL Viewer uses a binary compressed format (Macromolecular Transmission Format) to massively reduce network transfer and parsing times. These rich structural views of biological systems are provided to enable breakthroughs in scientific inquiry, medicine, drug discovery, technology, and education.

REPRODUCIBLE AND SCALABLE STRUCTURAL BIOINFORMATICS

Scientists face time-consuming barriers when applying structural bioinformatics analysis, including complex software setups, non-interoperable data formats, and lack of documentation – all which make it difficult to reproduce results and reuse software pipelines. A further challenge is the ever-growing size of datasets that need to be analyzed. To address these challenges, SDSC's Structural Bioinformatics Laboratory, directed by Peter Rose, is developing a suite of reusable, scalable software components called MMTF-PySpark, using three key technologies: its parallel distributed processing framework provides scalable computing; the MacroMolecular Transmission Format (MMTF), a new binary and compressed representation of Macromolecular structures, enables high-performance processing of Protein Data Bank structures; and Jupyter Notebooks is used to bundle code, 2D and 3D visualization, machine learning tools, and more into reproducible and reusable workflows. "The use of MMTF-PySpark could easily shave off a year of a graduate student's or postdoc's work in structural bioinformatics," said Rose. "We bank on contributions from the community to develop and share an ecosystem of interoperable tools."

MMTF.RCSB.ORG

IMPACT & INFLUENCE

STATE AND UC ENGAGEMENT

Aligning with Principles & Partnerships

A novel program that provides access to SDSC's supercomputing resources and expertise has to-date assisted some 40 individual research projects spanning all 10 UC campuses, with more than 18 million core hours of compute time allocated on SDSC's petascale *Comet* supercomputer funded by the National Science Foundation (NSF). Called HPC@UC and launched in mid-2016, the initiative is offered in partnership with the UC Vice Chancellors of Research and campus CIOs. To date, the program has helped researchers accelerate their time-to-discovery across a wide range of disciplines, from astrophysics and bioengineering to earth sciences and machine learning.

SDSC's HPC@UC program is specifically intended to:

- Broaden the base of UC researchers who require advanced and versatile computing;
- Seed promising computational research;
- Facilitate collaborations between SDSC and UC researchers;
- Give UC researchers access to cyberinfrastructure that complements what is available at their campus; and
- Help UC researchers successfully pursue larger allocation requests through NSF's eXtreme Science and Engineering Discovery Environment program (XSEDE), and other national computing programs.

WWW.SDSC.EDU/COLLABORATE/HPC_AT_UC.HTML

HIGH-PERFORMANCE COMPUTING WORKSHOPS

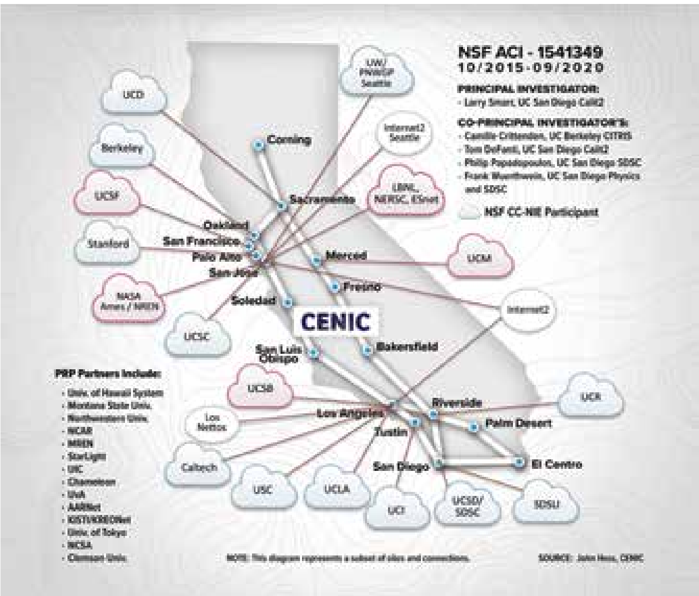
During the 2017/18 fiscal year, SDSC staff held numerous high-performance computing/'big data' workshops at various UC campuses and labs, including UC Berkeley, Lawrence Berkeley Lab, UCLA, and UC Santa Barbara, attracting hundreds of attendees that included graduate students, post-docs, and faculty. SDSC is now looking at ways to tailor similar courses to undergraduates. Course subjects included data analytics and machine learning using *Comet*, software tools for life sciences applications, as well as installing and using Python and Jupyter notebooks. SDSC's instructors hold Ph.Ds. in physics, astrophysics, aerospace engineering, computer science, cognitive science, and more. The workshops, which promote good interaction among UC researchers, have been conducted since the start of the UC@SDSC program in 2014.

SHERLOCK AND UCOP RISK MANAGEMENT

During the latest fiscal year, SDSC's Health CI Division, in partnership with the University of California's Office of the President (UCOP) Risk Services, deployed a secure, HIPAA-compliant, big data solution called Risk Services Data Management System within the Sherlock Cloud platform. "This secure platform provides a mechanism to collect, process, and transform UC Healthcare data from multiple sources and various formats into a single, integrated data set, enabling UCOP Risk Services enhanced management of UC's Liability Management initiative," said Sandeep Chandra, director of SDSC's Health CI Division. Announced in late 2017, the program is now fully operational following a successful launch. Read more about Sherlock on page 47.

PACIFIC RESEARCH PLATFORM

To meet the needs of researchers in California and beyond, the NSF awarded a five-year grant to fund the Pacific Research Platform (PRP). The PRP's data-sharing architecture, with end-to-end 10-100 gigabits per second (Gb/s) connections, will enable region-wide virtual co-location of data with computing resources and enhanced security options. Led by Calit2 Director Larry Smarr, Calit2 Researcher Tom DeFanti; SDSC's Frank Würthwein and Phil Papadopoulos; John Graham (UC San Diego); Camille Crittenden (UC Berkeley); John Hess (CENIC), Thomas Hutton (SDSC) and Eli Dart (ESnet); the PRP supports a broad range of data-intensive research projects that will have wide-reaching impacts on science and technology worldwide. Cancer genomics, human and microbiome 'omics' integration, biomolecular structure modeling, galaxy formation and evolution, telescope surveys, particle physics data analysis, simulations for earthquakes and natural disasters, climate modeling, virtual reality and ultra-resolution video development are just a few of the projects that are benefiting from the use of the PRP. The PRP will be extensible across other data-rich domains as well as other national and international networks, potentially leading to a national and eventually global data-intensive research cyberinfrastructure.



PRP.UCSD.EDU

From left to right: Frank Vernon, Graham Kent, John Graham, Ilkay Altintas, Patrick Perry (CIO, CSU), Louis Fox (President, CENIC). *Credit: DMT Imaging*

CENIC Recognizes WIFIRE and HPWREN as Top Innovations of 2018

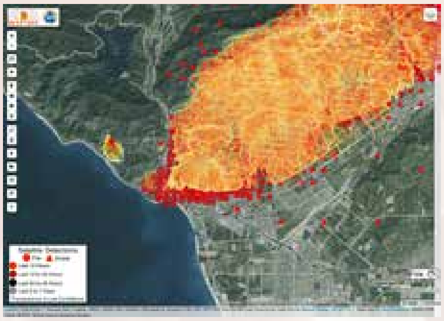
Two UC San Diego projects involving SDSC – WIFIRE and HPWREN (High Performance Wireless Research and Education Network) – were selected in early 2018 as recipients of the Corporation for Education Network Initiatives in California (CENIC) Innovations in Networking Awards for Experimental Applications, in recognition of work advancing technologies to help minimize damage and loss of life caused by wildfires. Awards went to SDSC Chief Data Science Officer and WIFIRE PI Ilkay Altintas and Frank Vernon (Scripps Institution of Oceanography), who now leads HPWREN after co-founding the organization with SDSC Research Scientist Hans Werner-Braun in 2000.

WIFIRE is an NSF-funded project that developed real-time and data-driven simulation, prediction, and visualization of wildfire behavior. During 2017's chaotic fires in Napa, Sonoma, Los Angeles, Ventura, and San Diego counties, WIFIRE's publicly available fire map was viewed more than 8 million times, while the WIFIRE team was in close communication with fire response agencies and chiefs from various fire departments. In December 2017, WIFIRE provided predictive maps for the Thomas, Skirball, Creek, Rye, and Lilac fires in Southern California and monitored the first responder radio channels and fire perimeter information to quickly create simulations of the spread of specific wildfires.

The collection of this crucial data was made possible by HPWREN, which has built high-speed wireless networks in San Diego, Imperial, Orange, and Riverside counties, enabling hundreds of cameras and meteorological stations to stream critically important data to servers connected to each other by the CENIC backbone, and providing wide-area wireless internet access throughout southernmost California.

CENIC also honored technology leaders with a new Founders Circle Award to recognize researchers who were instrumental in creating the network. Among those recipients were SDSC Network Architect Tom Hutton, and SDSC's Founding Director, Sidney Karin.

"Just as CENIC members today work collaboratively to shape and govern the organization, these leaders worked collaboratively to lay the groundwork necessary to launch the world-class research and education network used by millions of Californians," said CENIC President and CEO Louis Fox. Read more about HPWREN on page 18.



(Top) View from the HPWREN cameras atop Santiago Peak as the Holy Fire approaches in August 2018. *Image courtesy of HPWREN.*

(Bottom) The WIFIRE Firemap tool provided firefighters with predictive maps of the massive Thomas Fire as flames spread across Ventura County in December 2017. *Image courtesy of WIFIRE/SDSC.*

GOO.GL/GQWMTH

IMPACT & INFLUENCE

Image credit: Erik Jepsen, UC San Diego Publications

CAMPUS AND EDUCATION

STRENGTHENING TIES ACROSS CAMPUS AND OUR LOCAL COMMUNITIES

During the past fiscal year, SDSC closely aligned its investments with the priorities set forth in UC San Diego's recently approved strategic plan. As a result, SDSC has taken on a more integral role in key areas for campus, including research data management, data-enabled science, research computing, and education at the undergraduate/graduate/post-doc levels.

"Because we currently enjoy an overall robust financial health including operations, recharges, existing contracts and grants, and new funding sources, we are able to make significant strategic investments from our reserves, while realizing additional savings through ongoing operational efficiencies," SDSC Director Michael Norman told SDSC staff in spring 2018.

"With regard to our national cyberinfrastucture (CI) mission, we currently have 99 active awards totaling almost $150 million," said Norman. "Within our campus CI mission, we have a projected $3.9 million in service agreements including industry collaborations. And on the data science front, we have $1.6 million in revenue from our Health CI awards, and stand ready to support new research testbeds and at-scale projects in collaboration with the new Halicioğlu Data Science Institute (HDSI). Read more about SDSC and the HDSI on page 43.

"SDSC's development and operation of high-performance computing (HPC) resources at the national level provides substantial and tangible benefits to UC San Diego researchers, as well as the San Diego's burgeoning research communities," said Norman.

EMPOWERING THE NEXT GENERATION

SDSC's Education, Outreach, and Training (EOT) programs range from grade school to high school, helping students to become aware of opportunities within computational science at an early age, and then at the university level with numerous data science courses, including those in collaboration with HDSI. SDSC's initiative extends into serving the growing computational science workforce with workshops such as SDSC's Summer Institute, the International Conference on Computational Science, IEEE Women in Data Science Workshop, and more.

EDUCATION.SDSC.EDU

EDUCATIONAL IMPACT&INFLUENCE
ACROSS CAMPUS AND OUR LOCAL COMMUNITIES

ON THE LOCAL FRONT

RESEARCH EXPERIENCE FOR HIGH SCHOOL STUDENTS

The Research Experience for High School Students (REHS) program, a part of SDSC's student outreach program, was developed to help increase awareness of computational science and related fields of research to students in the San Diego region. Now in its ninth year, students gain exposure to career options, hands-on computational experience, and work readiness skills. During the eight-week summer program, students are paired with SDSC mentors to help them gain experience in an array of computational research areas. They learn how to formulate and test hypotheses, conduct computational experiments and draw conclusions from those experiments, and effectively communicate the science and societal value of their projects to a wide range of audiences. More than 400 students to-date have participated in REHS, and attendance levels have more than doubled in the last three years alone.

EDUCATION.SDSC.EDU/STUDENTTECH/?PAGE_ID=657

ON THE NATIONAL FRONT

ONLINE DATA SCIENCE & BIG DATA COURSES

UC San Diego recently launched a four-part Data Science series via edX's MicroMasters® program with instructors from UC San Diego's CSE Department and SDSC. In partnership with Coursera, SDSC created a series of MOOCs (massive open online courses) as part of a Big Data Specialization that has proven to be one of Coursera's most popular data course series. Consisting of five courses and a final Capstone project, this specialization provides valuable insight into the tools and systems used by big data scientists and engineers. In the final Capstone project, learners apply their acquired skills to a real-world big data problem. To date, the courses have reached more than 700,000 students in every populated continent – from Uruguay to the Ivory Coast to Bangladesh. A subset of students pay for a certificate of completion.

WWW.COURSERA.ORG/SPECIALIZATIONS/BIGDATA

HIGH-PERFORMANCE COMPUTING SUMMER INSTITUTE

SDSC's Summer Institute is an annual week-long training program offering introductory to intermediate topics on high-performance computing (HPC) and data science, with interactive classes and hands-on tutorials using SDSC's *Comet* supercomputer. The program has been expanded to cover new topics such as machine learning at scale, distributed programming in Python, cluster computing with Spark, and CUDA programming, while retaining traditional HPC topics such as performance tuning and parallel programming with MPI and OpenMP. The Summer Institute is aimed at researchers in academia and industry, especially in domains not traditionally engaged in supercomputing, and who may have challenges that cannot typically be solved using local computing resources. The 2018 event had more than 100 applicants from 54 institutions/companies. Generally, about half of the attendees are graduate students, with the remainder being post-docs, professors, and research staff at universities, national labs, and industry. SDSC has conducted the Summer Institute since the mid-1990s.

SI18.SDSC.EDU

The conference team (left to right): David Wang, The Bishop's School; Judy Qin, Canyon Crest Academy; Andrew (Junseong) Park, Canyon Crest Academy; Stephen Martindale, Torrey Pines High School; Susan Lin, Canyon Crest Academy; Adel Battikha, Westview High School; Jeremy Li, Del Norte High School; David Zeng, Westview High School; Ryan Li, Del Norte High School; Steven Chen, Westview High School; Valentina Kouznetsova, Ph.D. Associate Research Professor, UC San Diego/SDSC (mentor); Elliot Kim, San Dieguito Union High School

High School Students Present at Major Science Conference

Some 15 San Diego-area high school students who interned with SDSC presented research aimed at improving early diagnosis and treatment of a variety of ailments during the Third Annual Biomarkers International Conference, held last February in San Diego.

"We are pleased and proud of the participation of these young researchers in a national scientific conference of increasing importance to medical science," said Valentina Kouznetsova, an associate research professor with the Moores Cancer Center and SDSC.

"It's a rewarding experience to help these students gain traction toward a possible career in research," she added. Kouznetsova mentored the students in the biomarker program along with Igor Tsigelny, a research professor with the UC San Diego Department of Neurosciences and SDSC.

Calling High School Students for UC San Diego's Mentor Assistance Program

San Diego-area high school students interested in pursuing a career in scientifically-based research were invited to apply to UC San Diego's Mentor Assistance Program (MAP), a campus-wide initiative designed to engage students in a mentoring relationship with an expert from a vast array of disciplines. The latest mentoring period ran from September 2017 through May 2018.

Launched about two years ago by SDSC and UC San Diego School of Medicine, the MAP's mission is to provide a pathway for student researchers to gain access to UC San Diego faculty, post-doctoral fellows, Ph.D. candidates, and staff to mentor them in their own field of interest. Mentors are recruited from across campus from fields that include athletics, biology, chemistry, aerospace engineering, network architectures, pharmaceutical sciences, physics, social studies, and more.

"MAP is an opportunity for students to take the first step into a potential career path, while simultaneously building an early foundation for success in their academic career." said SDSC Education Manager and MAP co-founder Ange Mason. "These mentoring relationships are intended to support collegiality, effective communication, self-evaluation, and cultural competence, all of which enhance a stimulating and supportive university environment."

EDUCATION.SDSC.EDU/
STUDENTTECH/?PAGE_ID=2393

Hans-Werner Braun (top) is an SDSC research scientist who co-founded HPWREN with Frank Vernon (bottom), now its PI and a seismologist at the Scripps Institution of Oceanography.

A Cal Fire firefighter installs an antenna on a microwave tower to connect a nearby fire station to HPWREN. *Image courtesy of HPWREN/SDSC.*

HPWREN: A WIRELESS EDUCATION AND SAFETY NETWORK FOR GREATER SAN DIEGO

Initially funded by the National Science Foundation, the High Performance Wireless Research and Education Network (HPWREN) was co-founded by Hans-Werner Braun, a research scientist with SDSC; and Frank Vernon, a seismologist with the Scripps Institution of Oceanography who originally contacted Braun about creating a faster, more reliable, and more comprehensive network as new technology allowing real-time gathering of data.

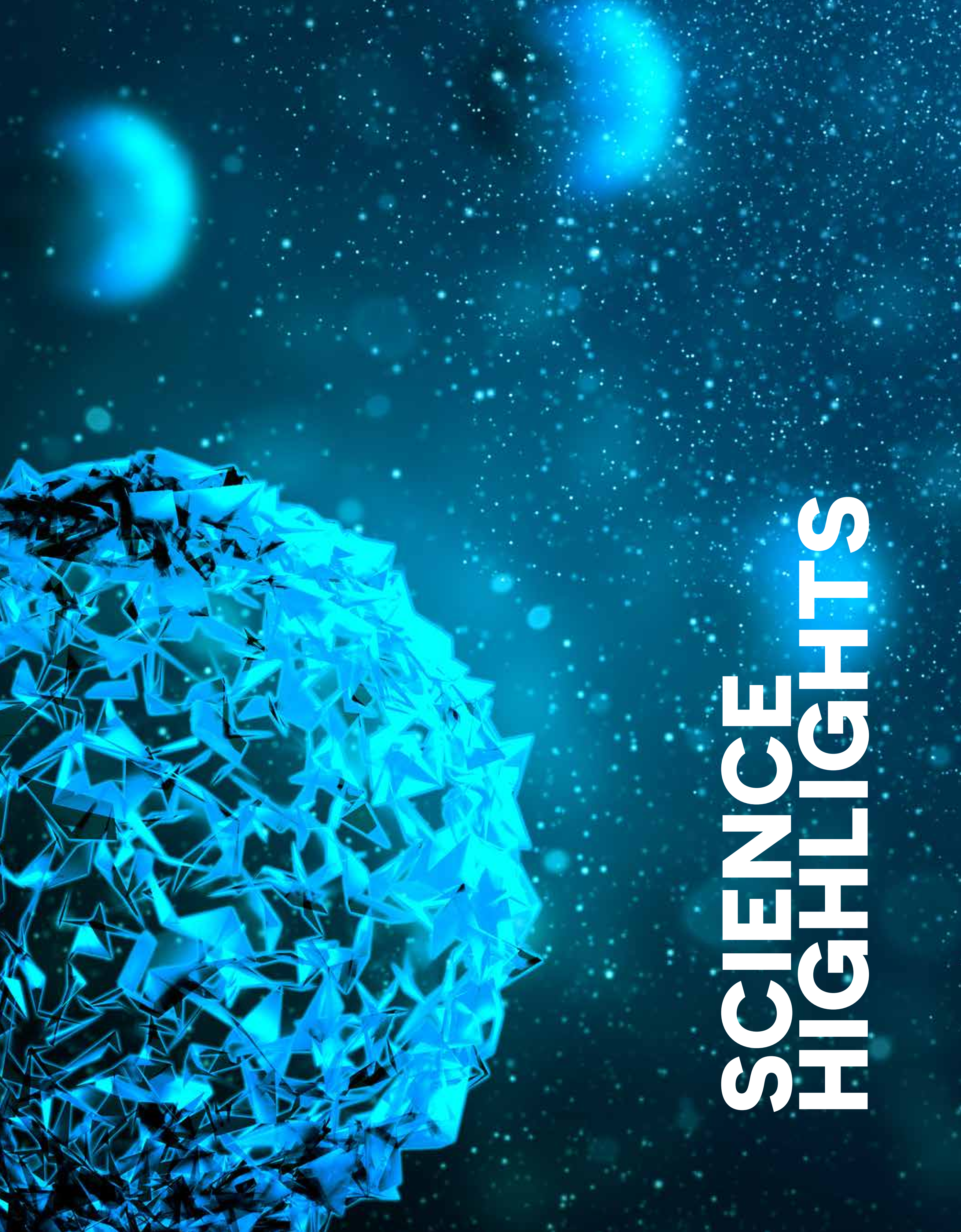
Today, HPWREN is an internet-connected "cyberinfrastructure" for research, education, and public safety that connects often hard-to-reach areas in remote environments via a system of cameras and weather stations to report local weather and environmental conditions, from severe rainstorms to wildfires and earthquakes.

In early 2018, the County of San Diego Board of Supervisors invested more than $437,000 for new cameras and a boost in the network speed of hazard detection technology developed by the HPWREN in partnership with the University of Nevada Reno ALERTWildfire project to improve the County's fire detection and response capabilities. "Our firefighters use this network daily to communicate among stations and monitor conditions in times of quiet and to fight fires when they arise," said San Diego Regional Fire Authority Chief Tony Mecham. "These improvements provide a valuable upgrade in the ability of our back country fire stations to have high-speed internet access and share information during both routine and emergency incidents."

HPWREN also supports the Area Situational Awareness for Public Safety Network (ASAPnet), an extension of the HPWREN infrastructure for the benefit of public safety communities, especially firefighters in San Diego County. ASAPnet consists of a wireless internet data communications overlay supporting rural fire stations and other firefighter assets, in addition to environment-observing cameras and other sensors.

Read more about HPWREN at the following QR code.

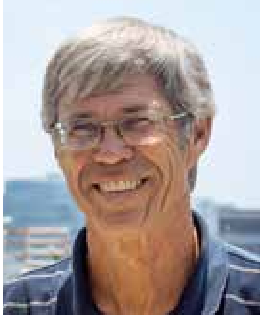
HPWREN.UCSD.EDU/
NEWS/20180215



SCIENCE HIGHLIGHTS

SCIENCE HIGHLIGHTS

SDSC Resources and Expertise Used in New Autism Study



SDSC Distinguished Scientist Wayne Pfeiffer.

SDSC Bioinformatics and Genomics Lead Madhusudan Gujral.

About a decade ago researchers established that gene mutations appearing for the first time, called *de novo* mutations, contribute to approximately one-third of cases of autism spectrum disorder (ASD). Now, a new study by an international team led by scientists at UC San Diego's School of Medicine has identified a culprit that may explain some of the remaining risk: rare inherited variants in regions of non-coding DNA mutations could come from fathers who aren't autistic.

The study, published in the April 20, 2018 issue of *Science*, shows that this culprit differs from known genetic causes in two ways. First, these variants don't alter the genes directly but instead disrupt the neighboring DNA control elements that turn genes on and off, called cis-regulatory elements or CREs. Second, these variants don't occur as new mutations in children with autism, but instead are inherited from their parents.

"For 10 years we've known that the genetic causes of autism consist partly of *de novo* mutations in the protein sequences of genes," said Jonathan Sebat, a professor of psychiatry, cellular and molecular medicine, and pediatrics at UC San Diego School of Medicine and chief of the Beyster Center for Genomics of Psychiatric Genomics. "However, gene sequences represent only 2 percent of the genome."

To investigate the other 98 percent of the genome in ASD, Sebat and his colleagues analyzed the complete genomes of 9,274 subjects from 2,600 families. One thousand were sequenced at Human Longevity Inc. (HLI) and at Illumina Inc., and DNA sequences were analyzed using SDSC's *Comet* supercomputer with the assistance of SDSC Distinguished Scientist Wayne Pfeiffer and SDSC Bioinformatics and Genomics Lead Madhusudan Gujral, who is also a co-author of the paper.

Using *Comet*, processing and identifying specific structural variants from a single genome took about 2½ days. "Since *Comet* has 1,984 compute nodes and several petabytes of scratch space for analysis, tens of genomes can be processed at the same time," said Pfeiffer. "Instead of months, with *Comet* we were able to complete the data processing in weeks."

The researchers then analyzed structural variants, deleted or duplicated segments of DNA that disrupt regulatory elements of genes, dubbed CRE-SVs. From the complete genomes of families, the researchers found that CRE-SVs that are inherited from parents also contributed to ASD. "We also found that CRE-SVs were inherited predominantly from fathers, which was a surprise," said co-first author William M. Brandler, a postdoctoral scholar in Sebat's lab at UC San Diego and a bioinformatics scientist at HLI.

SCIENCE.SCIENCEMAG.ORG/
CONTENT/360/6386/327.FULL

SCIENCE HIGHLIGHTS

EARTH SCIENCES

Earthquake Codes Developed by SDSC Used in 2017 Gordon Bell Prize Research

Senior Computational Scientist Yifeng Cui directs SDSC's High-Performance Geocomputing Laboratory.

A Chinese team of researchers that was awarded the prestigious Gordon Bell prize in late 2017 for simulating the devastating 1976 earthquake in Tangshan, China, used an open-source code developed by researchers at SDSC and San Diego State University (SDSU) with support from the Southern California Earthquake Center (SCEC).

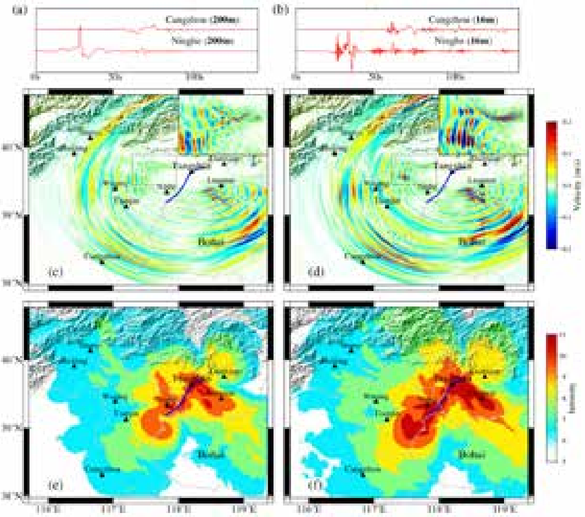
"We congratulate the researchers for their impressive innovations porting our earthquake software code, and in turn for advancing the overall state of seismic research that will have far-reaching benefits around the world," said Yifeng Cui, director of SDSC's High Performance Geocomputing Laboratory, who along with SDSU Geological Sciences Professor Kim Olsen, Professor Emeritus Steven Day, and researcher Daniel Roten developed the AWP-ODC code, which stands for 'Anelastic Wave Propagation, Olsen, Day, and Cui.'

The code is used by SCEC to simulate how earthquakes make the ground move, and is an important discovery for predicting ground motions that affect small homes and other structures, which are vulnerable to high-frequency shaking.

The 2017 Gordon Bell Prize – awarded each November by the Association for Computing Machinery (ACM) at the annual Supercomputing (SC) conference – went to the 12-member Chinese team for their research project called "18.9-Pflops Nonlinear Earthquake Simulation on *Sunway TaihuLight:* Enabling Depiction of 18-Hz and 8-Meter Scenarios."

Using the *Sunway TaihuLight*, which as of late 2017 was ranked as the world's fastest supercomputer, the team developed software that was able to efficiently process 18.9 Pflops (18.9 quadrillion calculations per second) of data and create three-dimensional visualizations of the 1976 earthquake that occurred in Tangshan and is believed to have caused between 240,000 and 700,000 casualties.

The team's software included innovations that achieved greater efficiency than had been previously attained running similar programs on the Oak Ridge National Laboratory's *Titan* supercomputer and the *TaihuLight* supercomputers in China.



Simulated ground motions of the 1976 Tangshan earthquake that resulted in a 12-member team of Chinese researchers being awarded the 2017 Gordon Bell prize.

Image courtesy of Haohuan Fu, Tsinghua University, et al.

GOO.GL/O3EDWG

SCIENCE HIGHLIGHTS

MATERIALS ENGINEERING

Making Bones Using Silky Secrets

SDSC's Comet Supercomputer Helps Sample Protein Folding in Bone Regeneration Study

Some secrets to repairing our skeletons might be found in the silky webs of spiders, according to recent experiments guided by supercomputers including SDSC's *Comet* system. The new study, which researchers say will help understand the details of osteoregeneration or how bones repair themselves, found that genes could be activated in human stem cells that initiate biomineralization, a key step in bone formation.

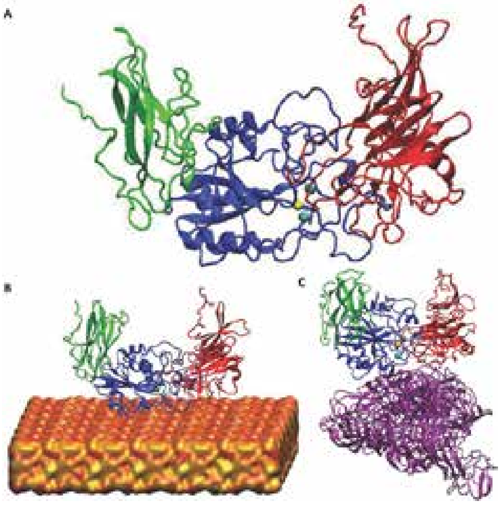
Scientists achieved these results with engineered silk derived from the dragline of golden orb weaver spider webs, which they combined with silica. The study, published in September 2017 in the journal *Advanced Functional Materials*, was the result of a combined effort from three institutions: Tufts University, Massachusetts Institute of Technology, and Nottingham Trent University.

Study authors used *Comet* and supercomputers at the Texas Advanced Supercomputing Center (TACC) via an allocation from the eXtreme Science and Engineering Discovery Environment (XSEDE) funded by the National Science Foundation (NSF). The supercomputers helped scientists model how the cell membrane protein receptor called integrin folds and activates the intracellular pathways that lead to bone formation. The research will help larger efforts to treat bone growth diseases such as osteoporosis or calcific aortic valve disease.

Silk has been shown to be a suitable scaffold for tissue regeneration due to its outstanding mechanical properties, explained study co-author Zaira Martín-Moldes a postdoctoral scholar at the Kaplan Lab at Tufts University who researches the development of new biomaterials based on silk. It's biodegradable. It's biocompatible. And it's fine-tunable through bioengineering modifications. The bone formation study targeted biomineralization, a critical process in materials biology.

"We would love to generate a model that helps us predict and modulate these responses both in terms of preventing the mineralization and also to promote it," said Martín-Moldes.

The researchers are building a pathway to generate biomaterials that could be used in the future, with the mineralization being a critical process. The final goal is to develop models that help design the biomaterials to optimize the bone regeneration process, when the bone needs to regenerate or to minimize it to reduce bone formation. These results help advance the research and are useful in larger efforts to help cure and treat bone diseases.



Simulated head piece domain of the integrin, based on molecular dynamics modeling. A) Integrin in solution. B) Integrin in contact with silica surface, modeled as having 4.7 silanol groups per nm2 and 0.45 siloxide groups per nm2. C) Integrin in contact with silk-chimera protein surface. Green: Hybrid domain of the β leg; Blue: βA domain of the β leg; Red: propeller domain of the β leg; Purple: silk-chimera protein; Orange: silica surface; Yellow: Mg2+ cation; Cyan: Ca2+ cation. Water molecules are not shown for clarity.

Image courtesy of Davoud Ebrahimi, MIT

goo.gl/7FS2Bk

SCIENCE HIGHLIGHTS

One-third of the Internet is Under Attack, According to SDSC Study

Alberto Dainotti is an assistant research scientist with the CAIDA group, based at SDSC.

For the first time, researchers have conducted a worldwide analysis of victims of internet denial-of-service (DoS) attacks.

Based on an analysis of activity from March 2015 to February 2017, researchers with SDSC's Center for Applied Internet Data Analysis (CAIDA) found that about one-third of the IPv4 address space was subject to some kind of DoS attacks, where a perpetrator maliciously disrupted services of a host connected to the internet. IPv4 is the fourth version of an Internet Protocol (IP) address, a numerical label assigned to each device participating in a computer network.

"We're talking about millions of attacks," said CAIDA researcher Alberto Dainotti, the report's principal investigator. "The results of this study are gigantic compared to what the big companies have been reporting to the public."

The study, presented in November 2017 at the Internet Measurement Conference in London and published in the *Proceedings of the Association for Computing Machinery* (IMC '17), sheds light on the types of DoS attacks as well as the adoption of commercial services to help combat these attacks.

"These results caught us by surprise in the sense that it wasn't something we expected to find. This is something we just didn't see coming," said Mattijs Jonker, the study's first author and a researcher with the University of Twente in The Netherlands who is a former CAIDA intern.

Two predominant types of DoS attacks, intended to overwhelm a service by a sheer mass of requests, are highlighted:

➢ "Direct" attacks, which involve traffic sent directly to the target from some infrastructure controlled by the attackers (e.g. their own machines, a set of servers, or even a botnet under their command.) These attacks often involve "random spoofing", characterized by faking the source IP address in the attack traffic.

➢ "Reflection" attacks, during which third-party servers are involuntarily used to reflect attack traffic toward its victim. Many protocols that allow for reflection also add amplification, causing the amount of reflected traffic sent toward the victim to be many times greater than that sent toward the reflector initially.

In a collaborative effort between UC San Diego, University of Twente, and Saarland University in Germany, researchers used two raw data sources that complement each other to detect the attacks: the UCSD Network Telescope, which captures evidence of DoS attacks that involve randomly and uniformly spoofed addresses; and the AmpPot DDoS (distributed denial-of-service) honeypots, which witness reflection and amplification of DoS attacks. Their data revealed more than 20 million DoS attacks that targeted about 2.2 million "slash 24 or /24" internet addresses (part of a routing number that denotes bit-length of the prefix), which is about one-third of the 6.5 million /24 blocks estimated to be alive on the internet.

"Put another way, during this recent two-year period under study, the internet was targeted by nearly 30,000 attacks per day," said Dainotti. "These absolute numbers are staggering, a thousand times bigger than other reports have shown."

Under a grant for the U.S. Department of Homeland Security (DHS), the CAIDA team plans to continuously monitor the DoS ecosystem to provide data for analysis to agencies and other researchers in a timely fashion.

GOO.GL/AJW8NY

EXPLORING THE UNIVERSE

High-Performance Computing's Role on Multi-Messenger Astronomy

Artist's illustration of two merging neutron stars. The rippling space-time grid represents gravitational waves that travel out from the collision, while the narrow beams show the bursts of gamma rays that are shot out just seconds after the gravitational waves. Swirling clouds of material ejected from the merging stars are also depicted. The clouds glow with visible and other wavelengths of light.

Image courtesy of NSF/LIGO/Sonoma State University/A. Simonnet

SCIENCE HIGHLIGHTS

Multi-messenger astronomy is based on the coordinated observation and interpretation of disparate signals, using ultra-sensitive devices such as high-powered telescopes, interferometers, particle detectors that pick up "messages" from electromagnetic radiation, gravitational waves, neutrinos, and cosmic rays. Such devices are now detecting billions of bits and bytes of information from colliding stars, merging galaxies, black holes and even the lingering cry from the birth of the Universe.

But how do researchers find nuggets of meaningful data — that could lead to landmark discoveries — from these mountains of information? Part of the solution lies in the use of high-performance computing (HPC), which can sift through all this data at lightning speeds and help researchers translate those bits and bytes into valuable insight.

"Advanced computing, along with experts charged with building and making the most of these HPC systems, has been critical to many Nobel Prizes, including work involving traditional modeling and simulation, to projects designed for more data-intensive workloads," said SDSC Director Michael Norman. As evidence, Norman and others point to several recent Nobel Prizes in chemistry and physics, including international collaborations exploring the dark side of the universe and others delving into the dynamics of proteins critical for tomorrow's targeted therapies.

Other HPC research requires the access, analysis, and interpretation of previously unfathomable amounts of data via a 'modality' called high-throughput computing (HTC) being generated from a wide cross-section of sensors and detectors. Simulation and data analysis along with experimentation sometimes complement and even blend with one another for discovery.

"HTC is a way of consuming computer resources, including those we label as HPC," said Frank Würthwein, professor of physics at UC San Diego and SDSC's Distributed High-Throughput Computing Lead. "The way these large-scale instruments do analysis requires the HTC 'modality' of computing. This is distinct from the standard 'submit a job to the queue' which is what people traditionally do for simulations."

SDSC's *Comet* supercomputer, a petascale system funded by the National Science Foundation (NSF) and capable of an overall peak performance of 2.6 petaflops or more than two quadrillion calculations per second, was one of several systems used to verify these recent major discoveries by cosmologists.

DISCOVERING COLLIDING NEUTRON STARS

The landmark discovery in late 2017 of gravitational and light waves generated by the collision of two neutron stars eons ago was in part made possible by a signal verification and analysis performed on SDSC's *Comet* supercomputer. The gravitational waves were detected using the NSF's Laser Interferometer Gravitational Wave Observatory (LIGO). Three LIGO-affiliated researchers won the 2017 Nobel Prize in Physics for an earlier detection in 2015 of gravitational waves in the universe as hypothesized by Albert Einstein some 100 years earlier.

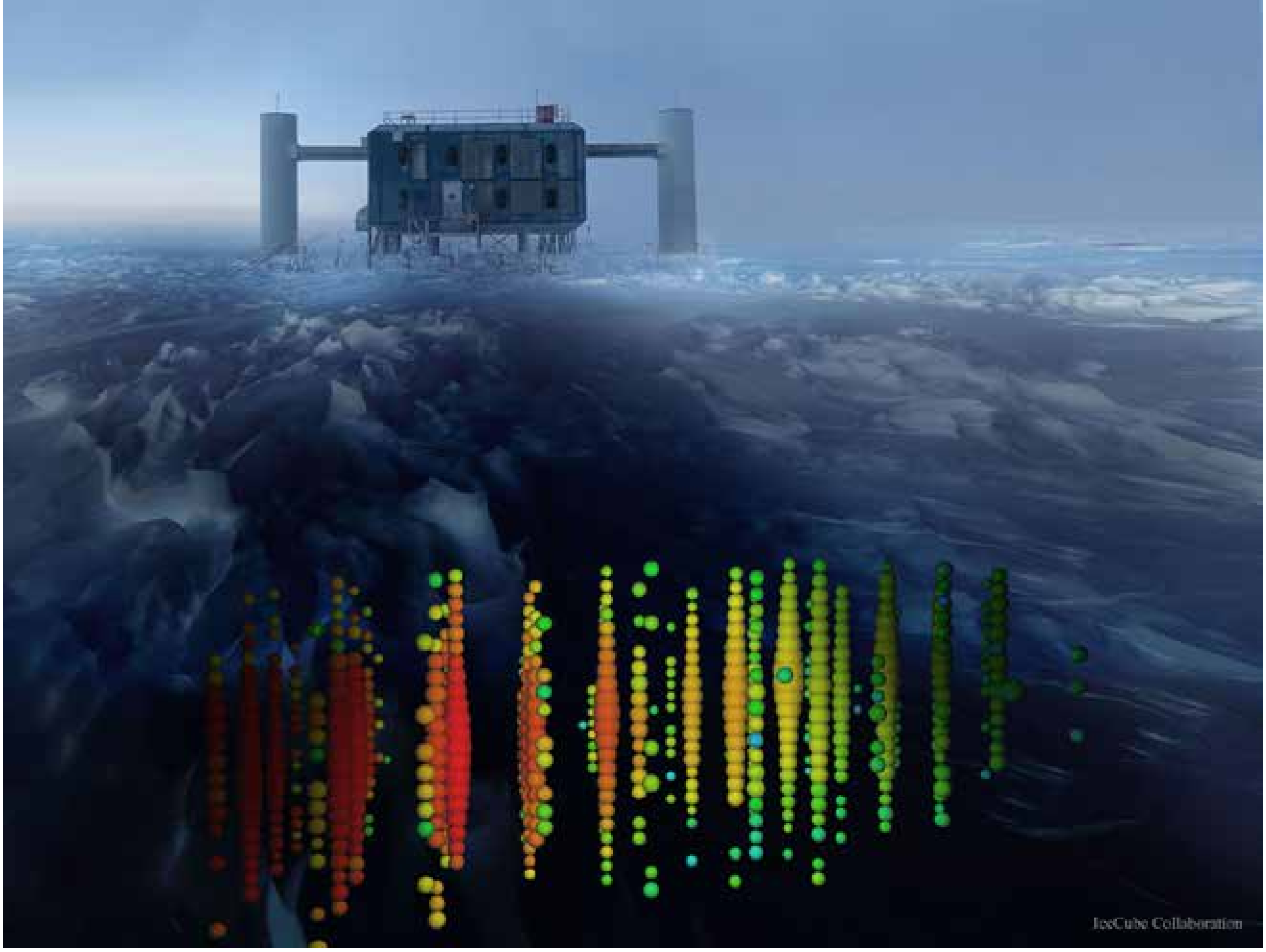
Comet was one of several supercomputers used to confirm the newest finding, as well as LIGO's initial discovery in 2015. As before, LIGO researchers benefited from high-throughput computing via *Comet* and the Open Science Grid (OSG), a multi-disciplinary research partnership specializing in large-scale computing services funded by the U.S. Department of Energy and the NSF. *Comet* reflects a commitment by SDSC to support high-throughput computing on national systems for large experimental facilities that started several years ago with the Center's support of large-scale data analysis for the Large Hadron Collider using SDSC's *Gordon* supercomputer.

LIGO researchers use many high-performance computing resources, but among those accessed via OSG and the NSF's eXtreme Science and Engineering Discovery Environment (XSEDE), *Comet* was the top resource used in terms of hours of computational time for both the 2015 and 2017 observations. According to Würthwein, LIGO researchers have so far consumed more than 2 million hours of computational time on *Comet* through OSG — including 600,000-900,000 hours each to help verify LIGO's findings in 2015 and the current neutron star collision — using *Comet's* Virtual Clusters for rapid, user-friendly analysis of extreme volumes of data.

Comet helped speed up signal verification and analysis of the event, allowing researchers to confirm their findings in a matter of days rather than weeks or months. "*Comet's* contribution through the OSG and XSEDE allowed us to rapidly turn around the offline analysis in about a day," said Duncan Brown, a LIGO collaborator and The Charles Brightman Professor of Physics at Syracuse University's Department of Physics. "That in turn allowed us to do several one-day runs, as opposed to having to spend several weeks before publishing our findings."

GOO.GL/BQUEP6

LIGO's gravitational wave detector in Livingston, Louisiana.
Image courtesy of Caltech/MIT/LIGO Lab



The IceCube Neutrino Observatory facility sits at the South Pole above an array of thousands of photodetectors. The sensors are deployed on strings of 60 modules each at depths between 1,450 to 2,450 meters (4,750 to 8,040 feet) into holes melted in the ice using a hot water drill. *Image courtesy of IceCube Collaboration, U. Wisconsin, National Science Foundation.*

FINDING FIRST EVIDENCE OF A COSMIC NEUTRINO SOURCE

In 1911 and 1912, Austrian physicist Victor Franz Hess made a series of ascents in a hydrogen balloon in a quest to find the source of ionizing radiation that registered on an electroscope. The prevailing theory was that the radiation came from the rocks of the Earth. During the last of his seven flights, Hess ascended to more than 5,300 meters – almost 17,400 feet – to find that the rate of ionization was three times greater than sea level. He concluded that the upper atmosphere is ionized by radiation from space, not the ground, and proved that this radiation is not solar after conducting experiments at night and during eclipses.

Hess had in fact discovered cosmic rays, and was awarded the Nobel Prize in Physics in 1936. Since the 1930s, cosmologists have been hunting for neutrinos – subatomic particles that can emerge from their sources and, like cosmological ghosts, pass through the universe unscathed, traveling for billions of light years from the most extreme environments in the universe to Earth.

In September 2017, an international team of scientists identified a source of high-energy cosmic neutrinos for the first time. "We have never before used multi-messenger astrophysics to pinpoint the origin of high-energy cosmic rays," said NSF Director France Córdova in announcing the finding.

The detection was made by scientists at the NSF-funded IceCube Neutrino Observatory, an array of 5,160 optical sensors deep within a cubic kilometer of ice at the South Pole. The findings were confirmed by telescopes around the globe and in Earth's orbit. The scientific importance of this observation lies in the fact that the source for such a high-energy cosmic ray could be convincingly identified as a known blazar, roughly four billion light years from Earth. A blazar is a giant elliptical galaxy with a massive, rapidly spinning black hole at its core.

Once again, *Comet* was used for discovery, this time to answer "one of the oldest open questions in astronomy" according to Francis Halzen, a University of Wisconsin–Madison professor of physics and the lead scientist for the IceCube Neutrino Observatory.

NEUTRINOS
goo.gl/QSVNXw

MULTI-MESSENGER ASTRONOMY
youtu.be/nZTB8NxwPTQ

FOCUSED SOLUTIONS& APPLICATIONS

FOCUSED SOLUTIONS& APPLICATIONS

FOR **ADVANCED COMPUTATION**

SDSC's computational, storage, and networking resources – along with a high level of expertise required to configure, operate, and support them – create an advanced cyberinfrastructure to support scientific discovery across numerous disciplines spanning academia, industry, and government.

Advanced but user-friendly resources such as SDSC's petascale-level *Comet* supercomputer underscore a vital need for systems that serve a broad range of research, with a focus on researchers who have modest-scale computational needs, which is where the bulk of computational science needs exist.

Comet has established itself as one of the most widely used supercomputers in the National Science Foundation's (NSF) eXtreme Science and Engineering Discovery Environment (XSEDE) program, which connects researchers to an advanced collection of integrated digital resources and services.

"SDSC's national mission to help pioneer an advanced research cyberinfrastructure has always been our core, and that has enabled us to support collaborations at the local and state levels," said SDSC Director Michael Norman, who is also the principal investigator for the *Comet* program, the result of an NSF grant now totaling more than $27 million.

COMET

Cyberinfrastructure for the Long Tail of Science

Within its first 18 months of operation, SDSC's *Comet* supercomputer soared past its initial goal of serving 10,000 unique users across a diverse range of science disciplines, from astrophysics to phylogenetics.

By mid-2018, more than 33,000 individual researchers used *Comet* to run science gateways jobs alone since the supercomputer entered production in mid-2015. A science gateway is a community-developed set of tools, applications, and data services and collections that are integrated through a web-based portal or suite of applications. Another 4,700 users from over 350 institutions accessed *Comet* via traditional login. Since entering operations, *Comet* has provided more than 900 million compute core-hours and 5M GPU hours of computing across a wide range of science disciplines.

While *Comet* is capable of an overall peak performance of 2.6 petaflops – or 2.6 quadrillion calculations per second – its allocation and operational policies are geared toward rapid access, quick turnaround, and an overall focus on scientific productivity.

In mid-2018 the NSF extended *Comet's* service into a sixth year of operation, with the system now slated to run through March 2021. The research community depends on long-term availability and continuity in computing resources, and this $2.4M supplemental award from NSF ensures continued access to this highly productive and user-friendly system.

GPU Count Doubled

SDSC recently doubled the number of graphic processing units (GPUs) on *Comet* in direct response to growing demand for GPU computing among a wide range of research domains. The expansion makes *Comet* the largest provider of GPU resources available to the NSF's XSEDE program. Under a supplemental NSF award valued at just over $900,000, SDSC expanded *Comet* with the addition of 36 GPU nodes, each with four NVIDIA P100s, for a total of 144 GPUs. This doubles the number of GPUs from the previous 144 to 288 and represents the largest GPU resource available through the XSEDE program. Applications include but are not limited to GPU-memory management systems such as VAST, analysis of data from large scientific instruments, and molecular dynamics software packages such as AMBER, LAMMPS, and BEAST – the latter used extensively by SDSC's Cyberinfrastructure for Phylogenetic Research (CIPRES) science gateway, which receives the majority of its computing resources from *Comet*.

Triton Shared Computing Cluster (TSCC)

High-Performance Computing for Researchers and Innovators

The *Triton Shared Computing Cluster (TSCC)*, operated on behalf of UC San Diego by SDSC, continued to deliver high-performance research computing capabilities for campus and industry investigators. *TSCC* is a "condo computing" program established in 2013 that has continued to grow in popularity among campus and external users during the FY2017-18 period. Condo computing is a research computing best practice, wherein researchers use grant or startup funds to purchase and contribute computer servers ("nodes") to the cluster, building up a significant computing resource. In return for their contribution, researchers get access to the full cluster on a "fair share" basis.

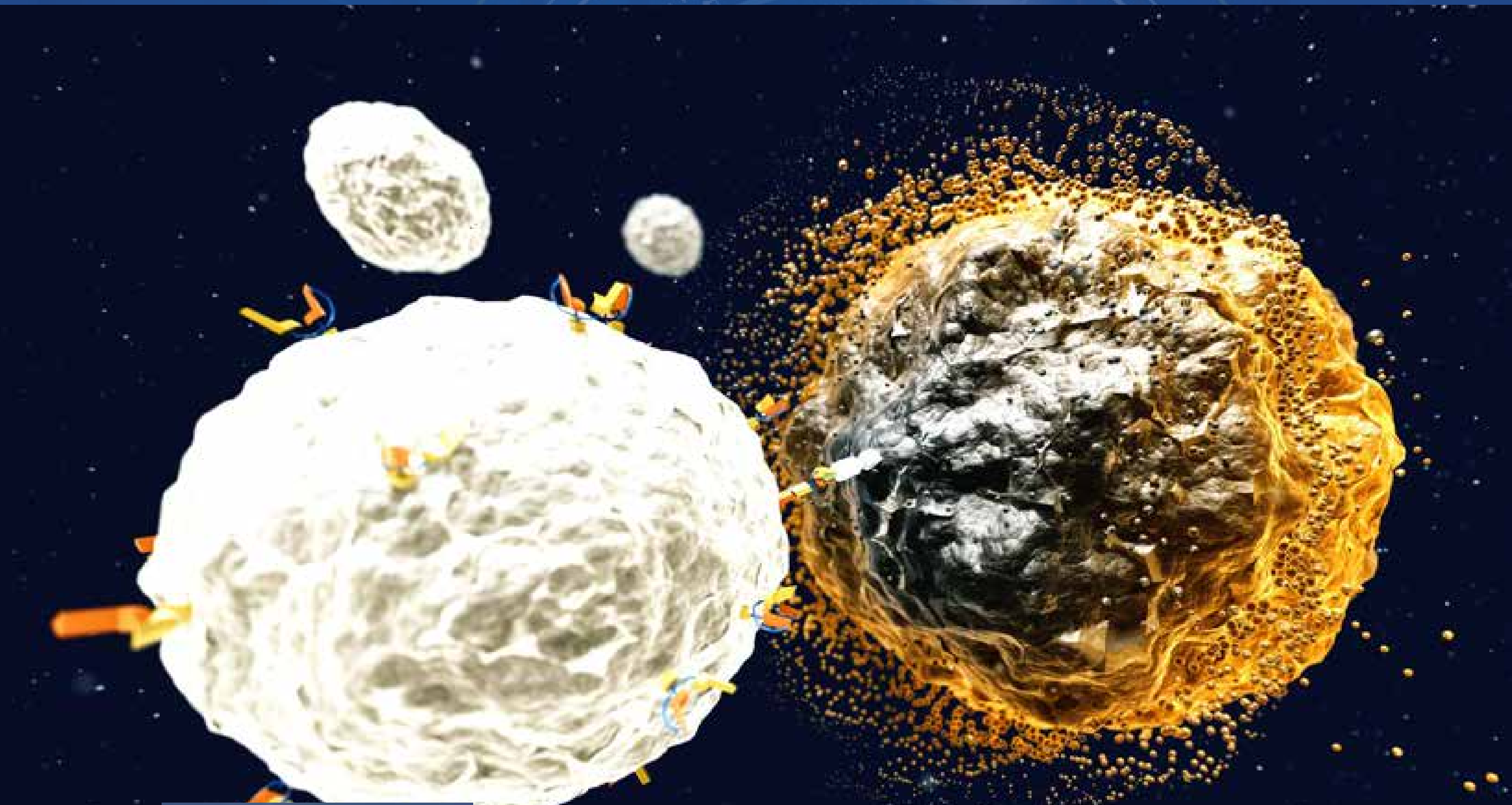
Today, *TSCC* is supporting research across a wide variety of domains, including biology and life sciences, chemistry, climate, engineering, political and social sciences, and others. Since its launch, *TSCC* has grown to 34 participating labs/groups with almost 300 researcher-owned compute nodes, plus an additional 75 common nodes available to anyone on campus through a pay-as-you-go recharge model. The latter is popular with individual researchers with occasional or temporary computing needs, students, and classes needing access to high-performance computing.

The 2017-2018 period was marked by two significant developments – welcoming a major new participant to the cluster, UC San Diego Assistant Professor Ludmil Alexandrov; and full implementation of the 'BioBurst' cluster addition to *TSCC.*

Alexandrov joined UC San Diego's Bioengineering Department in 2017 after serving as an Oppenheimer Fellow in the Theoretical Biology and Biophysics Group and the Center for Nonlinear Studies at Los Alamos National Laboratory. His research (see facing page) relies heavily on computational methods and required a significant augmentation of *TSCC* in terms of computing power and storage. On Alexandrov's behalf, SDSC purchased and installed 36 new compute nodes equipped with the latest Intel Skylake high-performance processors from Dell Corporation. SDSC also purchased one petabyte of parallel file system storage and 200 terabytes of durable storage to support Alexandrov's needs.

Mid-2018 marked the full implementation of SDSC's BioBurst cluster project, an upgrade to *TSCC* to deliver targeted capabilities for bioinformatics analyses. The NSF grant, valued at almost a half-million dollars, is part of the NSF's Campus Cyberinfrastructure (CC*) program, which invests in coordinated campus-level cyberinfrastructure (CI) components of data, networking, computing infrastructure, capabilities, and integrated services. The project provided funding to expand *TSCC's* capacity and deploy new technology designed specifically to accelerate the analysis of DNA, RNA, and related bioinformatics data.

GOO.GL/WWAQL5



Ludmil Alexandrov is an assistant professor with UC San Diego's Cellular & Molecular Medicine Department, and an Oppenheimer Fellow in the Theoretical Biology & Biophysics Group and the Center for Nonlinear Studies at Los Alamos National Laboratory.

Pursuing the Fight against Cancer

Professor Ludmil Alexandrov's research has been focused on understanding mutational processes in human cancer through the use of mutational signatures. In 2013, he developed the first comprehensive map of the signatures of the mutational processes that cause somatic mutations in human cancer. This work was published in several well-regarded scientific journals and highlighted by the American Society of Clinical Oncology as a milestone in the fight against cancer. More recently, Alexandrov mapped the signatures of clock-like mutational processes operative in normal somatic cells, demonstrated that mutational signatures have the potential to be used for targeted cancer therapy, and identified the mutational signatures associated with tobacco smoking.

Alexandrov has 55 publications in peer-reviewed journals including 14 publications in *Nature*, *Science*, or *Cell* and another 17 publications in *Nature Genetics*, *Nature Medicine*, or *Nature Communications*. In 2014, he was recognized by Forbes magazine as one of the "30 Brightest Stars under the Age of 30" in the field of Science & Healthcare. In 2015, he was awarded the Prize for Young Scientists in Genomics and Proteomics by *Science* magazine and *SciLifeLab*, and also received a Harold M. Weintraub Award by the Fred Hutchinson Cancer Center. In 2016, he was awarded the Carcinogenesis Young Investigator Award by Oxford University Press and the European Association for Cancer Research. Alexandrov is currently one of six co-investigators leading the Mutographs of Cancer project, a $25 million initiative that seeks to fill in the missing gaps to identify the unknown cancer-causing factors and reveal how they lead to cancer.

SDSC Co-location (Colo) Facility

SDSC offers colocation services to UC San Diego, the UC system, and the local research community. SDSC's 19,000 square-foot climate-controlled, secure data center is equipped with 13 megawatts of power, multi- 10, 40, and 100 gigabit network connectivity, and a 24/7 operations staff. Within that center is a "co-location" facility that is free to UC San Diego researchers via a program aimed at saving campus funds by housing equipment in an energy-efficient, shared facility. The Colo facility houses computing and networking equipment for dozens of campus departments, every division and school, as well as local partners that include Rady Children's Hospital, the J. Craig Venter Institute, Simons Foundation, The Scripps Research Institute, and the Sanford-Burnham Medical Research Institute.

SDSC's Colo facility has resulted in more than $2.1 in million annual energy savings and about $1.5 million in avoided construction costs in the latest fiscal year, while streamlining and improving the management of hundreds of campus systems. The facility is well-suited to installations that need to demonstrate regulatory compliance, as well those that require high-speed networking. SDSC welcomes inquiries from local companies interested in co-locating equipment to facilitate collaborations with UC San Diego and SDSC investigators.

GOO.GL/MXIMIJ

Storage, Networking & Connectivity

Large-scale Academic Deployment of Cloud Storage and Compute

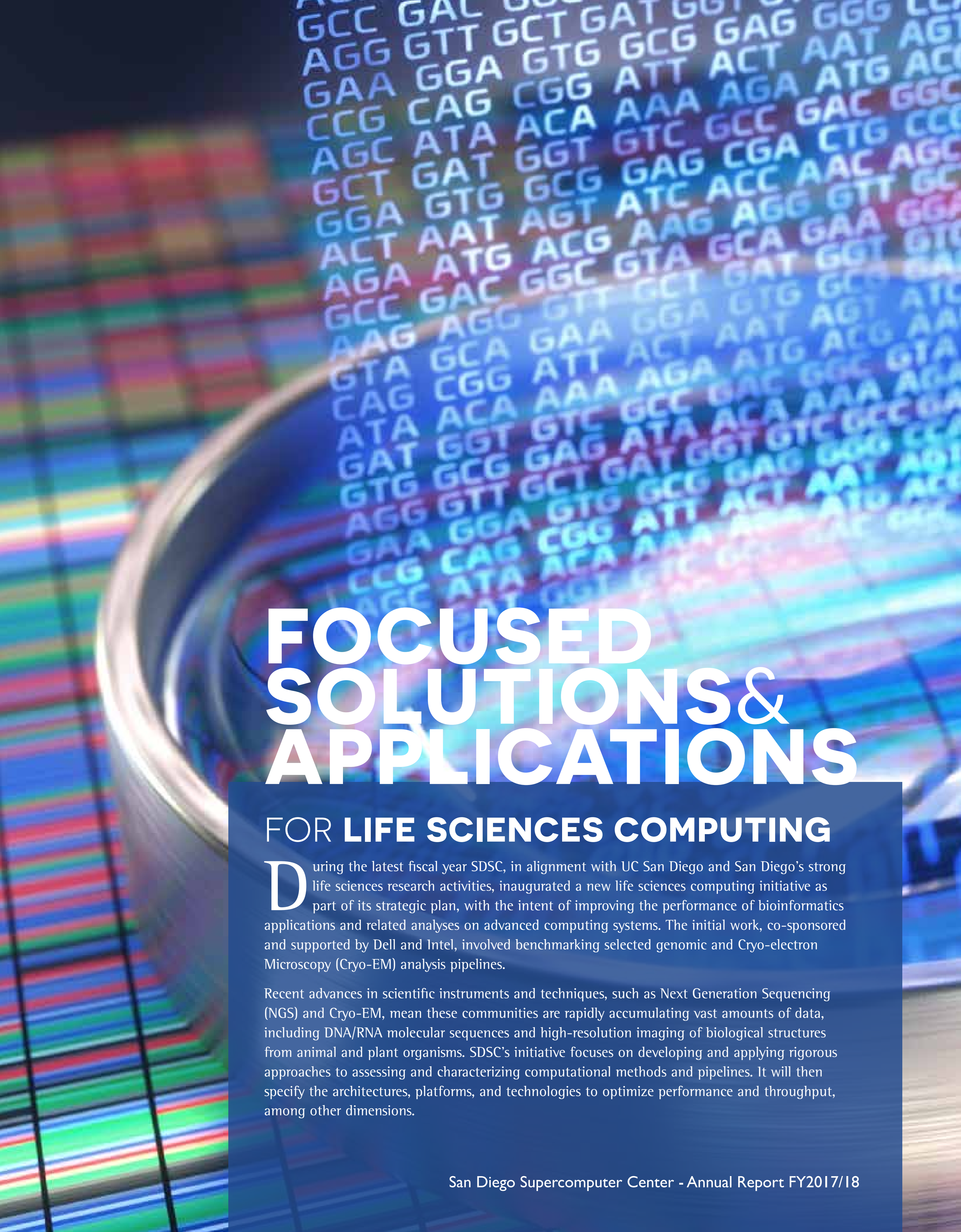
SDSC's Research Data Services team administers a large-scale storage and compute cloud. UC San Diego campus users, members of the UC community, and UC affiliates are eligible to join the hundreds of users who already benefit from the multi-petabyte, OpenStack Swift object store. SDSC Cloud boasts a simplified recharge plan that eliminates fees such as bandwidth and egress charges. SDSC Cloud also includes an elastic compute facility, based on OpenStack Nova, using Ceph for storage. This comprehensive cloud environment provides researchers with a testbed and development environment for developing cloud-based services and for many data science workflows. It is especially attuned to data sharing, data management, and data analytics services. SDSC Cloud is one of the platforms that underpins the National Data Service's suite of offerings (see page 7 for more on NDS). UC San Diego researchers who make use of SDSC Cloud using sponsored research funds will no longer incur overhead charges.

Data Oasis

SDSC's *Data Oasis* is a Lustre-based parallel file storage system that provides high performance I/O to *Comet* and *TSCC*. A critical component of SDSC's Big Data initiatives, *Data Oasis* features 12 petabytes (PB) of capacity and 200 gigabytes (GB) per second of bandwidth. *Data Oasis* ranks among the fastest parallel file systems in the academic community. Its sustained speeds mean researchers could retrieve or store 240 terabytes (TB) of data—the equivalent of *Comet's* entire DRAM memory—in about 20 minutes, significantly reducing time needed for retrieving, analyzing, storing, or sharing extremely large datasets. Recent enhancements to Lustre, such as Data on Metadata, improvements to the *Data Oasis* storage networking, and the presence of flash memory on HPC system flash memory on SDSC's HPC systems, combine to give users a flexible, high-performance storage environment capable of tackling the most demanding simulation and data analysis problems.

SDSC Universal Scale Storage

In March 2018, SDSC launched a scale-out storage service to UC San Diego, the UC system, and the local research community. To meet the needs of researchers, SDSC built a universal scale storage service with Qumulo, a modern approach to scale-out storage that delivers fast and flexible storage with the real-time data usage and performance analytics necessary for visibility at the petabyte scale, all while containing costs. This service creates a single namespace that can be used in research labs or on SDSC's supercomputers and clusters. Universal scale storage already provides multiple petabytes of space to its current users and continues to rapidly expand to meet demand.

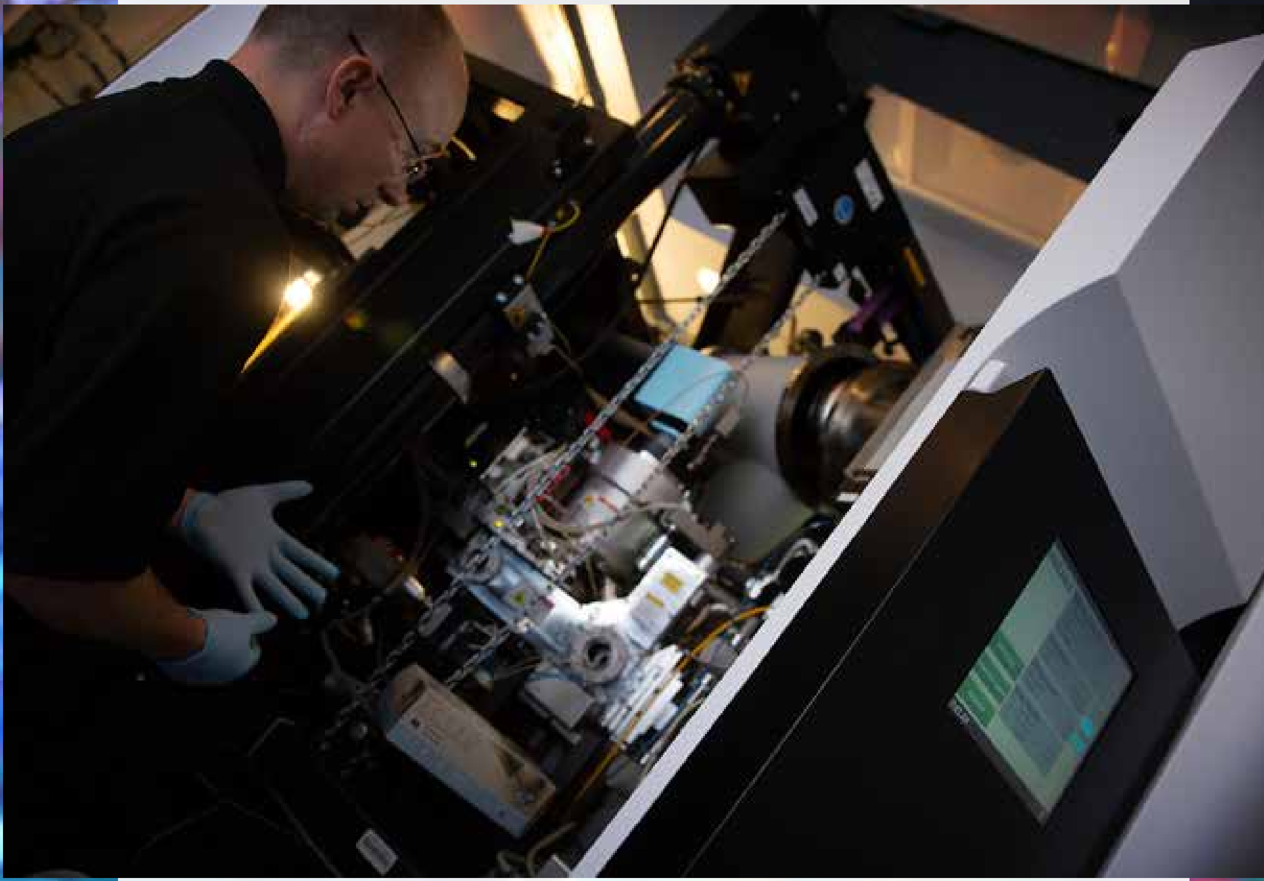


FOCUSED SOLUTIONS & APPLICATIONS

FOR **LIFE SCIENCES COMPUTING**

During the latest fiscal year SDSC, in alignment with UC San Diego and San Diego's strong life sciences research activities, inaugurated a new life sciences computing initiative as part of its strategic plan, with the intent of improving the performance of bioinformatics applications and related analyses on advanced computing systems. The initial work, co-sponsored and supported by Dell and Intel, involved benchmarking selected genomic and Cryo-electron Microscopy (Cryo-EM) analysis pipelines.

Recent advances in scientific instruments and techniques, such as Next Generation Sequencing (NGS) and Cryo-EM, mean these communities are rapidly accumulating vast amounts of data, including DNA/RNA molecular sequences and high-resolution imaging of biological structures from animal and plant organisms. SDSC's initiative focuses on developing and applying rigorous approaches to assessing and characterizing computational methods and pipelines. It will then specify the architectures, platforms, and technologies to optimize performance and throughput, among other dimensions.



In cryo-Electron Microscopy (cryo-EM), biological samples are flash-frozen so rapidly that damaging ice crystals are unable to form. As a result, researchers are able to view highly-detailed reconstructed 3D models of intricate, microscopic biological structures in near-native states. Above is a look inside of one of the cryo-electron microscopes available to researchers at the Timothy Baker Lab at UC San Diego. *Image credit: Jon Chi Lou, SDSC.*

In the past few decades, the life sciences have witnessed one landmark discovery after another with the aid of high-performance computing, paving the way toward a new era of personalized treatments based on an individual's genetic makeup, and drugs capable of attacking previously intractable ailments with few side effects. Genomics research is generating torrents of biological data to help develop personalized treatments, believed to be the focus for tomorrow's medicine. The sequencing of DNA has rapidly moved from the analysis of data sets that were megabytes (millions) in size to entire genomes that are gigabytes (billions) in size. Meanwhile, the cost of sequencing has dropped from about $10,000 per genome in 2010 to $1,000 in 2017, thus requiring increased speed and refinement of computational resources to process and analyze all this data.

In one of the most extensive genome analyses performed at SDSC, an international team led by Jonathan Sebat, a professor of psychiatry, cellular and molecular medicine and pediatrics at UC San Diego School of Medicine, recently identified a risk factor that may explain some of the genetic causes for autism: rare inherited variants in regions of non-code DNA.

For about a decade, researchers knew that the genetic cause of autism partly consisted of so-called de novo mutations, or gene mutations that appear for the first time. But those sequences represented only 2 percent of the genome. To investigate the remaining 98 percent of the genome in ASD (autism spectrum disorder), Sebat and colleagues analyzed the complete genomes of 9,274 subjects from 2,600 families, representing a combined data total on the range of terabytes (trillions) of bytes. Read more about this study on page 21.

CIPRES Awarded Two Federal Grants to Support Innovations in Biological Research

More than $2.8 million provided by NSF and NIH for Popular Phylogenetic Gateway

CIPRES, a gateway to major discoveries about the genetic relationships of our planet's living creatures, was awarded more than $2.8 million in grants from the National Science Foundation (NSF) and the National Institutes of Health (NIH) during the latest fiscal year that together will expand its software and resource capabilities for biological research.

Short-hand for CyberInfrastructure for Phylogenetic RESearch, CIPRES is a web-based gateway launched at SDSC in 2009 that allows researchers to explore evolutionary connections among species using supercomputers provided by the NSF's eXtreme Science and Engineering Discovery Environment (XSEDE) project. The gateway provides access to a sophisticated set of software tools and high-performance computers that would be both costly and difficult for individual researchers to create.

Last year, CIPRES supported more than 10,000 researchers who are investigating a wide range of biological fields, including the evolutionary history of proteins, viruses, bacteria, plants, and animals; identifying new genera and species; evaluating and developing new techniques; and exploring the evolutionary history of diverse populations. At least 4,500 peer-reviewed publications in journals have relied on CIPRES resources for their research, with results appearing in some of biology's most prestigious journals such as *Science*, *Nature*, *Proceedings of the National Academy of Sciences* (PNAS), and *Cell*.

"Understanding the evolutionary history of living organisms is a central goal of nearly every discipline in biology," Mark Miller, a bioinformatics researcher at SDSC and the gateway's principal investigator. "We're gratified by CIPRES' wide acceptance among the biological community, and pleased that both federal agencies have seen fit to support its future operations."

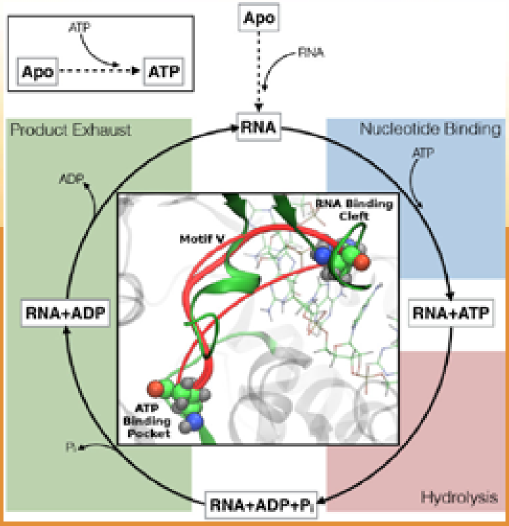
The NIH award provides about $1.86 million over four years to update CIPRES' infrastructure. Specifically, users have suggested incorporating advances in web technologies, such as graphic interface tools that are now integral parts of smart phones and tablets, in addition to analytical software tools that could speed discoveries.

"Our goal is to provide tools to increase the rate of discovery in any field that relies on high-throughput data acquisition, and where the rate of discovery is limited by the rate of data analysis," said Wayne Pfeiffer, an SDSC Distinguished Scientist and the project's co-principal investigator.

YOUTU.BE/GN4MJWZAI1K

The NSF award provides about $1 million over three years, providing support to sustain the project's resources while it seeks funds elsewhere for innovation. During this time, the project's staff is expected to seek extramural funding to improve the gateway's usability, job efficiency, job restarts, access to cloud computing, and data sharing and publication. "The funding under this program will help to ensure that CIPRES remains a stable, reliable resource for the broad biology community for the next three years," said Miller.

Supercomputer Simulations Reveal New "Achilles heel" in Dengue Virus



Dengue virus NS3 helicase utilizes energy from ATP binding and hydrolysis to power the unwinding and translocation along viral RNA. Microsecond long molecular dynamics trajectories have identified motif V as communication hub between the ATP pocket and RNA binding cleft. These findings suggest that motif V is a new target for the development of selective antivirals.

Image courtesy of Russell Davidson and Martin McCullagh, Colorado State University

In a recent edition of *PLOS Computational Biology*, the simulations, aided by the use of SDSC's *Comet* supercomputer, home in on a small segment of a viral enzyme called non-structural protein 3 (NS3), that plays a critical role in replicating the dengue RNA genome, which the virus requires to survive and spread.

During the past couple of decades, researchers have considered NS3 a primary target for developing drugs capable of inhibiting and preventing the replication of the dengue virus. But many worry that since NS3's protein sequence has similarities to related human proteins, drugs capable of inhibiting this enzyme might create unwanted side effects, even affecting a cell's natural antiviral response. For this reason, researchers have been trying to further clarify how this viral enzyme works on a molecular scale. By stretching the amount of time proteins can be simulated in their natural state of wiggling and gyrating, a team of researchers at Colorado State University has identified a critical protein structure that could serve as a molecular Achilles heel, able to inhibit the replication of dengue virus and potentially other flaviviruses such as West Nile and Zika virus.

"We hope that some of these findings will be common amongst other flaviviruses," said Martin McCullagh, assistant professor of chemistry at Colorado State University and the study's principal investigator. "We're currently working on a follow-up study on the Zika NS3 protein as well as continuing our collaboration with Brian Geiss to test some of our findings using experimental biochemical assays and virology."

Additional co-authors of this study include Russell B. Davidson and Josie Hendrix, both from Colorado State University.

GOO.GL/SVV85H

FOCUSED SOLUTIONS& APPLICATIONS

FOR DATA-DRIVEN PLATFORMS AND APPLICATIONS

SDSC's mission has expanded to encompass more than advanced computation, as researchers require innovative applications and expertise related to the ever-increasing amount of digitally-based scientific data. The Center's current strategic plan focuses on three main areas: Advanced and versatile computing, data science and engineering, and life sciences computing.

"For more than 30 years – and as one of the first four U.S. supercomputer centers opened in 1985 by the National Science Foundation – SDSC has long been at the national forefront of large-scale scientific computing, which is foundational to the emergence of data-enabled scientific research," said SDSC Director Norman. "SDSC is a living laboratory for students seeking a career in scientific research, as well as high-tech companies and other organizations that are seeking partnerships to advance their research through the use of high-performance and data-intensive computing."

UC San Diego Alumnus Taner Halıcıoğlu speaks at the March 2, 2018 dedication event for the Halicioğlu Data Science Institute, co-located at SDSC. *Image credit Jon Chi Lou, SDSC*

Halicioğlu Data Science Institute & SDSC's Data Science Hub

UC San Diego's official launch of the Halicioğlu Data Science Institute (HDSI) in March 2018 was welcomed by SDSC, which will house research labs and offices for HDSI's senior staff and faculty in the Center's East Building. "We're pleased that SDSC will be home to many HDSI-affiliated faculty and staff as this exciting new initiative gets underway," said SDSC Director Michael Norman. "We look forward to SDSC being a hub of connectivity for data analytics and innovation for the entire UC San Diego campus."

UC San Diego Chancellor Pradeep K. Khosla led the dedication event at SDSC's auditorium, following an announcement in 2017 of a $75 million donation from UC San Diego Alumnus Taner Halicioğlu (hah-li-jyo-loo) to create the innovative institute. The gift from Halicioğlu was the largest ever received from a UC San Diego alumnus. Halicioğlu currently is a private investor as well as a lecturer in UC San Diego's Department of Computer Science and Engineering.

HDSI.UCSD.EDU

HDSI, established as an academic unit of UC San Diego, is a campus-wide initiative that focuses on cross-disciplinary education, research and industry outreach involving computer science, cognitive science, mathematics, and other fields. In keeping with the Halicioğlu Institute mission, affiliated researchers and staff will also work in other campus locations, such as Qualcomm Institute at Atkinson Hall, to help interweave data science efforts into the greater UC San Diego community

"It just made sense to have the institute at SDSC because this place is all about data," said Halicioğlu, who interned at SDSC 1995-96 while earning his bachelor's degree in computer science. In remarks during the event, he noted, "Data science has technically always existed, but I don't think it really started to coalesce into an actual discipline until recently. I'm excited that we'll hopefully be on the forefront of that."

DATASCIENCE.SDSC.EDU

The new Halicioğlu Institute has been collaborating with other research units at UC San Diego, including SDSC's Data Science Hub, where experts at SDSC and other parts of the university can apply their experience in building multi-disciplinary data science teams to help provide solutions to regional, national, and global challenges such as smart cities, precision medicine, advanced manufacturing, and data center automation. "We very much look forward to working with HDSI by not only developing new applications for data-driven research and analytics, but by actively training and thereby establishing a modern data science workforce that can help drive innovation," said Ilkay Altintas, SDSC's chief data science officer and director of the SDSC Data Science Hub.

As the co-founder and director of HDSI, Rajesh Gupta, a professor in the Department of Computer Science and Engineering, looks forward to working with the campus community and industry. "The HDSI foundation includes a campus cyberinfrastructure established by SDSC that will become one of the core resources for HDSI programs."

'AWESOME' Social Media Data Platform

Amarnath Gupta, leader of the AWESOME project, is director of the Advanced Query Processing Lab and a full research scientist at SDSC.

SDSC researchers have developed an integrative data analytics platform that harnesses the latest 'big data' technologies to collect, analyze, and understand social media activity, along with current events data and domain knowledge. The platform has the capability to continuously ingest multiple sources of real-time social media data and scalable analysis of such data for applications in social science, digital epidemiology, and internet behavior analysis.

Called AWESOME (Analytical Workbench for Exploring SOcial MEdia), the platform is assisting social science researchers, global health professionals, and government analysts by using real-time, multi-lingual, citizen-level social media data and automatically crosslinking it to relevant knowledge to better understand the impact on and reaction to significant social issues.

Funded by the National Science Foundation (NSF) and National Institutes of Health (NIH), AWESOME's goal is to benefit society through areas as diverse as detecting free speech suppression, to shaping policy decisions or even slowing the spread of viruses. "Our tagline is 'heterogeneous analytics for heterogeneous data'," said SDSC researcher Amarnath Gupta, the project lead.

After being awarded a NIH grant shared by SDSC, UCLA, and UC Irvine, AWESOME is being used to make timely predictions for the number of high HIV-risk patients at county levels. The award is the result of a collaboration between the groups under the banner of the the University of California's Office of the President (UCOP)-supported UC Institute for Prediction Technology (UCIPT), a multi-campus program to use new innovations in social technologies to predict human behaviors and outcomes.

"The goal is to figure it out if we can find not people who have HIV but people who might be at risk for HIV by looking at their social media behavior," said Gupta, noting that UC San Diego has a mobile clinic for HIV testing, and the goal is to move the unit to locations where more people have the opportunity to get tested. "At no point is the individual's data tracked, and there is no action taken on the individual at all."

A second award, which also started in September 2017, was granted to UC San Diego and Princeton University from the NSF RIDIR (Resource Implementations for Data Intensive Research) program to perform multi-lingual integrative analysis of textual data from multiple sources such as newspapers, Twitter, Weibo, political biographies, etc., with the goal of identifying emerging situations that may require the attention of policy-makers and crisis managers.

"We talk a lot about 'smart' cities, but I believe we should also talk about smart citizens who are more aware of the world around them," said Gupta. "(They) can interact and express their opinion and voices more clearly so they need to be informed... for example if a political situation is unraveling or an epidemic outbreak is happening."

youtu.be/dQk4Hkr5RJY

AWESOME uses AsterixDB, a scalable data management system that can store, index, and manage semi-structured data that resulted from a research project at UC Irvine and UC Riverside. SDSC is collaboratively extending AsterixDB to enable new political and social science analytics for the 21st Century China Center initiative of UC San Diego, a project that produces scholarly research and informs policy discussions on China and U.S.-China relations.

SDSC Centers of Excellence

SDSC's Centers of Excellence are part of a broad initiative to assist researchers across many data-intensive science domains, including those who are relatively new to computational and data-enabled science. These centers represent key elements of SDSC's wide range of expertise, from 'big data' management to the analysis and advancement of the internet.

SDSC Chief Data Science Officer Ilkay Altintas also directs the WorDS Center and is principal investigator for the WIFIRE project, a university-wide collaboration funded by the National Science Foundation (NSF) to create a cyberinfrastructure to effectively monitor, predict, and mitigate wildfires.

WORKFLOWS FOR DATA SCIENCE (WorDS) CENTER

Called WorDS for 'Workflows for Data Science', this center of excellence combines more than a decade of experience within SDSC's Scientific Workflow Automation Technologies laboratory, which developed and validated scientific workflows for researchers working in computational science, data science, and engineering. "Our goal with WorDS is to help researchers create their own workflows to better manage the tremendous amount of data being generated in so many scientific disciplines, while letting them focus on their specific areas of research instead of having to solve workflow issues and other computational challenges as their data analysis progresses from task to task," said Ilkay Altintas, director of the center and SDSC's Chief Data Science Officer.

Funded by a combination of sponsored agreements and recharge services. WorDS' expertise and services include:

- World-class researchers and developers well-versed in data science, big data, and scientific computing technologies;
- Research on workflow management technologies that resulted in the collaborative development of the popular Kepler Scientific Workflow System;
- Development of data science workflow applications through a combination of tools, technologies, and best practices;
- Hands-on consulting on workflow technologies for big data and cloud systems, i.e., MapReduce, Hadoop, Yarn, Spark, and Flink; and
- Technology briefings and classes on end-to-end support for data science.

WORDS.SDSC.EDU

Sandeep Chandra is director of SDSC's Health Cyberinfrastructure (CI) and head of the Sherlock center of excellence.

SDSC SHERLOCK CLOUD

Sherlock, an offering of SDSC's Health Cyberinfrastructure Division (Health CI Division), is focused on secure and compliant, compute, data management and big data services for academia, government and industry partners. Sherlock Cloud, a multi-tenant, scalable hybrid Cloud, is compliant with the Federal Information System Management Act (FISMA), Health Information Portability and Accountability Act (HIPAA), and NIST 800-171 Controlled Unclassified Information (CUI) requirements. Sherlock Cloud successfully collaborated with Amazon Web Services (AWS) to duplicate its managed services in the AWS Cloud, giving customers the option to choose managed compliant services operating on-site at SDSC or in the AWS Cloud.

During the latest fiscal year SDSC's Health CI Division, in partnership with the UCOP's Risk Services, deployed a secure, HIPAA-compliant, big data solution called Risk Services Data Management System within the Sherlock Cloud platform. "This secure platform provides a mechanism to collect, process, and transform UC Healthcare data from multiple sources and various formats into a single, integrated data set, enabling UCOP Risk Services enhanced management of UC's Liability Management initiative," said Sandeep Chandra, director of SDSC's Health CI Division. Announced in late 2017, the program is now fully operational following a successful launch.

SHERLOCK.SDSC.EDU

CAIDA's founder and director, KC Claffy, is a resident research scientist at SDSC whose research interests span internet topology, routing, security, economics, future internet architectures, and policy.

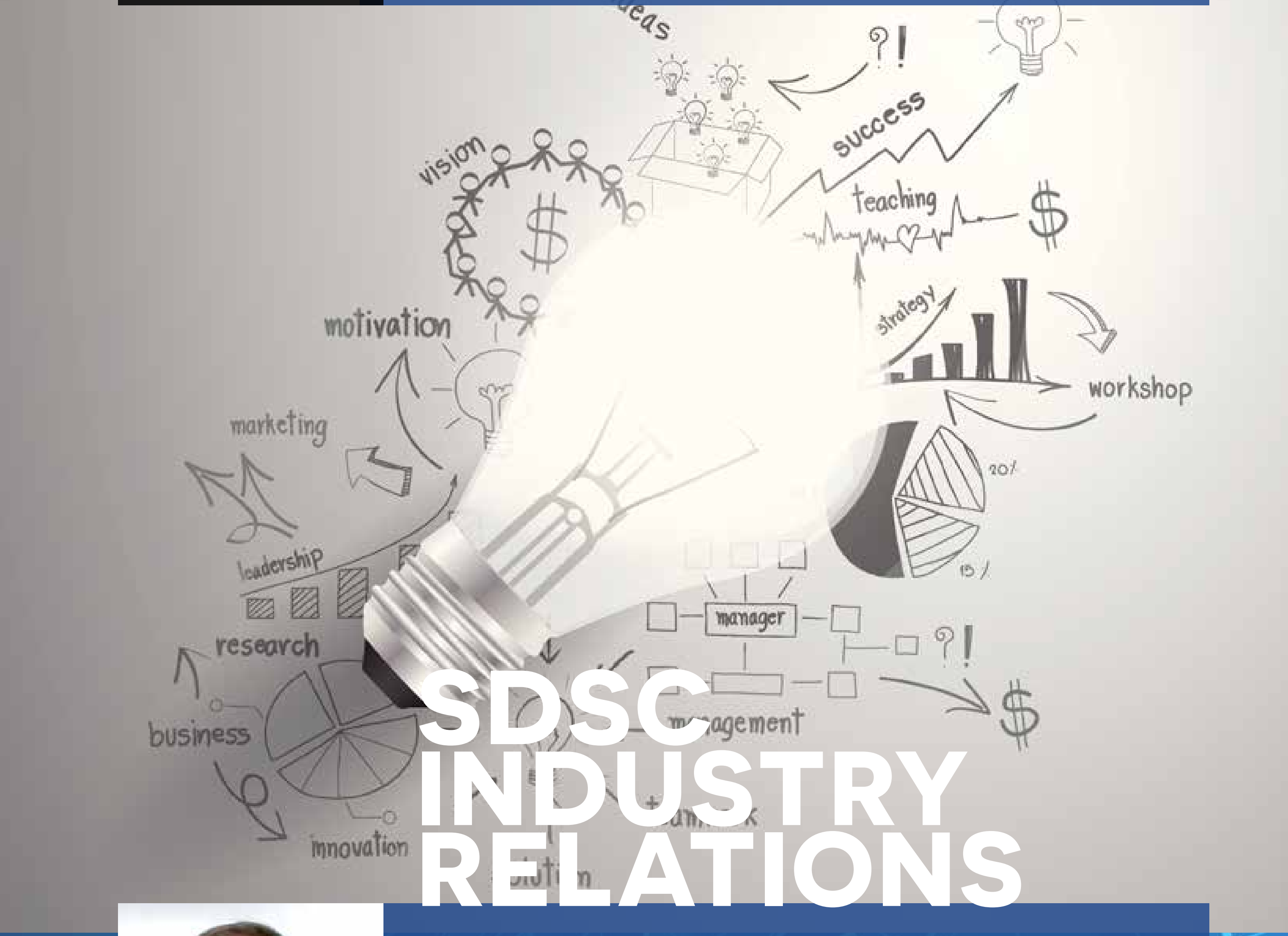
CAIDA

Internet Research for Cybersecurity and Sustainability

The Center for Applied Internet Data Analysis (CAIDA) was the first center of excellence at SDSC. Formed in 1997, CAIDA is a commercial, government, and research collaboration aimed at promoting the engineering and maintenance of a robust and scalable global internet infrastructure. CAIDA's founder and director, KC Claffy, is a resident research scientist at SDSC whose research interests span internet topology, routing, security, economics, future internet architectures, and policy.

CAIDA was recently awarded a $4 million, five-year NSF grant to integrate several of its existing research infrastructure measurement and analysis components into a new Platform for Applied Network Data Analysis, or PANDA. The platform, to include a science gateway component, is in response to feedback from the research community that current modes of collaboration do not scale to the burgeoning interest in the scientific study of the internet. Read more about CAIDA and KC Claffy on Pages 4 and 27.

CAIDA.ORG



SDSC INDUSTRY RELATIONS

Ron Hawkins is director of Industry Relations for SDSC and manages the Industry Partners Program, which provides member companies with a framework for interacting with SDSC researchers and staff to develop collaborations.

HELPING DRIVE THE INNOVATION ECONOMY

As a research university that in recent years has attracted over $1 billion in research funding annually, UC San Diego is naturally a hub for new ideas and innovation. The interplay of a major research university with an entrepreneurial and collaborative private sector makes San Diego a nexus for startup companies and new ventures, contributing to both the local and national economies. As the largest 'Organized Research Unit' at UC San Diego, SDSC is a key attractor and facilitator for early-stage companies, especially in advanced computing technology and data science. A snapshot in time in mid-2017 revealed no less than nine startup/early-stage companies working with SDSC, ranging from precision medicine for oncology to software-defined networking, to a next-generation DNA sequencing platform, to a non-von Neumann processor architecture. SDSC's Industry Partners Program provides a portfolio of mechanisms to support new ventures, from spinout or licensing of technologies, to technical evaluation and benchmarking, provision of computing and storage services, standup of pilot/proof-of-concept/demonstration projects, training programs, and others. SDSC values the opportunity to engage and work with local innovators. Going forward, SDSC seeks to be a part of and contribute to one of the most vibrant, entrepreneurial localities globally.

INDUSTRY PARTNERS PROGRAM

In early 2018 SDSC launched a revamped Industry Partners Program (IPP), beginning with a new series of activities aimed at fostering industrial research engagements, including quarterly breakfast roundtables covering a range of technology topics relevant to both science and commerce. The new IPP serves as a gateway for companies that would like to further develop or augment their expertise in specific parts of today's diverse high-tech economy. The "IPP 2.0" structure offers multiple levels of engagement opportunities through its Technology Forum, Supporter/Partner research collaborations, SDSC's Advanced Technology Lab, HPC and data science training, and other activities.

The Technology Forum is an annual, fee-based program that offers member companies access to SDSC researchers via breakfast roundtable conversations and other activities covering the latest technologies and research in high-performance computing, data science, and related areas. Recent topics covered include HPC and storage trends in life sciences and a distributed, national-scale data caching platform for 'big data'. The events attracted existing and prospective Forum and IPP members from around San Diego and beyond. A wide-ranging series of topics is planned for 2018-2019, in addition to an SDSC Industry Day and research review at the annual Data West conference in December 2018.

SDSC'S ADVANCED TECHNOLOGY LAB CONTINUES TO SUPPORT TECHNOLOGY DEVELOPMENT

Through SDSC's Advanced Technology Lab (ATL), researchers are evaluating new computing hardware and software technologies to better understand how they can be used in future computing systems for the national research community, and to assist participating companies in evaluating and advancing their technologies and products. Working closely with the private sector and other organizations, ATL researchers gain early access to the newest technologies, in many cases before they are generally available or while still under development.

During this reporting period, researchers in the ATL engaged with numerous industry partners:

- **Engility Corporation:** ATL staff and their students applied data mining methods and machine learning algorithms to analyze I/O traffic of users' jobs on high-performance computers (HPC).
- **Department of Defense, Engility:** To analyze asynchronous parallel I/O performance of multi/many-core HPC resources and develop an I/O Library which makes it easier to implement parallel I/O for DoD scientists.
- **GigaIO:** ATL staff evaluated GigaIO's new high-performance interconnect fabric for HPC and machine learning/artificial intelligence applications. The fabric permits a disaggregation of GPU accelerators and SSD storage from compute nodes via a direct PCIe switching fabric, creating a more flexible and cost-effective arrangement for HPC and ML/AI.
- **Memcomputing:** To evaluate its unique non-von Neumann computing architecture for solving complex optimization problems.
- **WekaIO:** ATL formed a partnership to evaluate a "hyper-converged" flash file system on SDSC's *Comet* supercomputer and *TSCC* campus research cluster.

INDUSTRY.SDSC.EDU

WWW.DATAWEST.ORG

ATL.SDSC.EDU

Facts & Figures

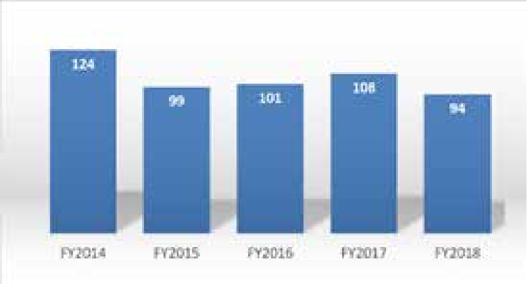
Proposal Success Rate

	FY14	FY15	FY16	FY17	FY18
Proposals Submitted	77	91	73	84	75
Proposals Funded	39	38	33	33	26
Success Rate	50%	42%	45%	39%	35%

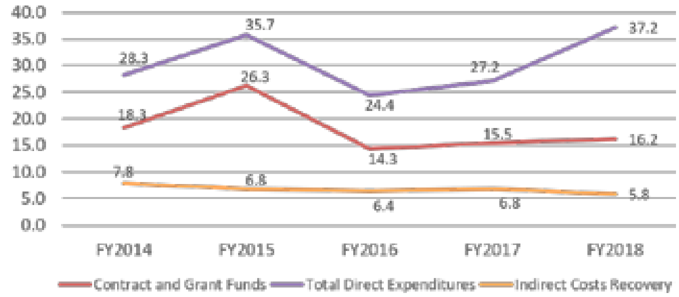
(Source: Campus Data Report)

In perhaps the most competitive landscape for federal funding in the last two decades, SDSC's overall success rate on federal proposals averages 42% over the last five years compared to the FY2017 national average of about 18% for computer science and engineering proposals at the National Science Foundation.

Number of Sponsored Research Awards

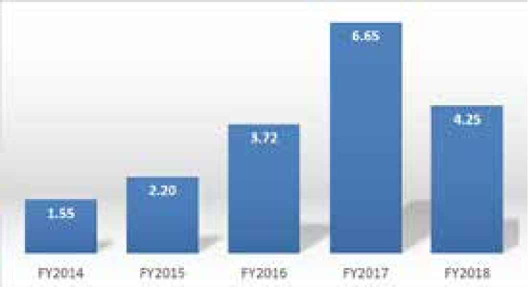


Total Expenditures ($M)



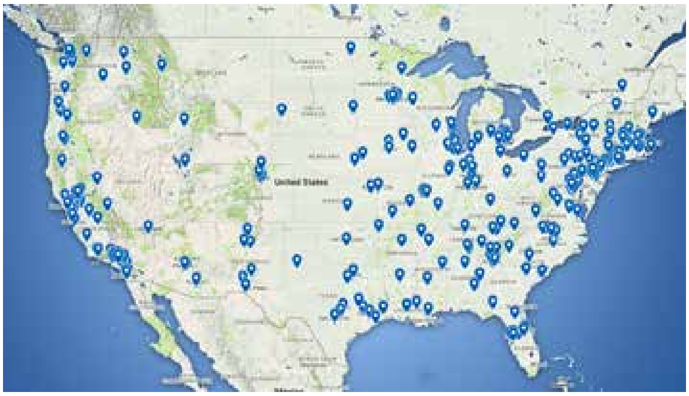
Apart from the extraordinary research impact of SDSC collaborations and partnerships, a quick look at the fiscal impact of these collaborations is impressive. During its 32-year history, SDSC revenues have exceeded $1 billion, a level of sustained funding matched by few academic research units in the country. At the close of FY2018, SDSC had 48 NSF-funded projects totaling $90 million.

Total Revenue from Industry ($M)



In 2017, the jump in revenue reflects a $2M contract lasting two years. The FY2018 bar in the chart above does not reflect this revenue since it was counted in FY2017.

Geographical Distribution of National Users of SDSC HPC Resources



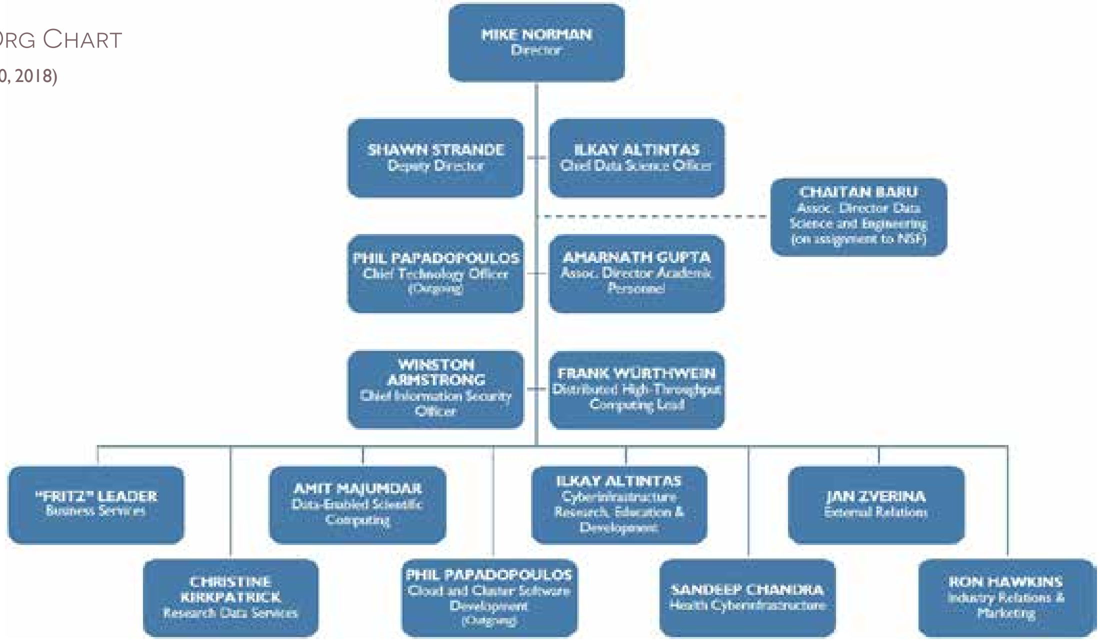
A total of 2,907 unique users from around the world (17 unique countries) accessed SDSC's *Comet* HPC resource directly. Of these users, 2,769 were based in the United States. The map to the left illustrates a sampling of the U.S. locations of these users. In addition, more than 26,780 unique users conducted research on *Comet* via science gateways.

On *Comet*, a total of 320,193,966 *Comet* core hours and 2,022,532 GPU hours were used during the FY2017 period.

Organization & Leadership

SDSC Org Chart

(as of June 30, 2018)


MIKE NORMAN
Director
SHAWN STRANDE
Deputy Director
ILKAY ALTINTAS
Chief Data Science Officer
CHAITAN BARU
Assoc. Director Data Science and Engineering (on assignment to NSF)
PHIL PAPADOPOULOS
Chief Technology Officer (Outgoing)
AMARNATH GUPTA
Assoc. Director Academic Personnel
WINSTON ARMSTRONG
Chief Information Security Officer
FRANK WÜRTHWEIN
Distributed High-Throughput Computing Lead
"FRITZ" LEADER
Business Services
AMIT MAJUMDAR
Data-Enabled Scientific Computing
ILKAY ALTINTAS
Cyberinfrastructure Research, Education & Development
JAN ZVERINA
External Relations
CHRISTINE KIRKPATRICK
Research Data Services
PHIL PAPADOPOULOS
Cloud and Cluster Software Development (Outgoing)
SANDEEP CHANDRA
Health Cyberinfrastructure
RON HAWKINS
Industry Relations & Marketing

Executive Team

Ilkay Altintas
Chief Data Science Officer

Chaitanya Baru (on assignment to NSF)
Associate Director, Data Science and Engineering

Sandeep Chandra
Division Director, Health Cyberinfrastructure

Ronald Hawkins
Director, Industry Relations

Christine Kirkpatrick
Division Director, Research Data Services

Samuel "Fritz" Leader
Chief Administrative Officer

Amit Majumdar
Division Director, Data-Enabled Scientific Computing

Michael Norman
SDSC Director

Philip M. Papadopoulos (outgoing)
Division Director, Cloud & Cluster Software Development

Shawn Strande
Deputy Director

Frank Würthwein
Lead, Distributed High-Throughput Computing

Jan Zverina
Division Director, External Relations

UC External Advisory Board

Kimberly S. Budil
UC, Office of the President (UCOP)

Michael Carey
UC Irvine

Paul Dodd
UC Davis

Adams Dudley
UC San Francisco

Lise Getoor
UC Santa Cruz

Ralph Greenspan
UC San Diego

Eamonn Keogh
UC Riverside

Juan C. Meza
UC Merced

Peter Nugent
Lawrence Berkeley National Laboratory and UC Berkeley

John Sarrao
Los Alamos National Laboratory

Tim Sherwood
UC Santa Barbara

Paul Weiss
UCLA

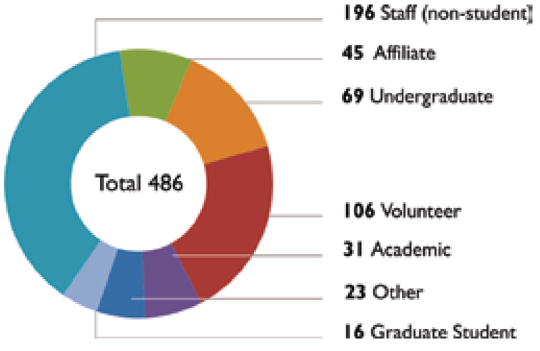
Tarek I. Zohdi
UC Berkeley

Executive Committee

UC San Diego

Sandra Brown
Mark Ellisman
Michael Holst
J. Andrew McCammon
John Orcutt
Al Pisano (chair)
Tajana Rosing
Nicholas Schork
Brian Schottlaender (outgoing)
Robert Sullivan
Susan Taylor
Gabriel Wienhausen

SDSC Census



RESEARCH EXPERTS
SDSC Computational & Data Scientists

Ilkay Altintas, Ph.D
Chief Data Science Officer, SDSC
Director, Workflows for Data Science (WorDS) Center of Excellence
Lecturer, Computer Science and Engineering, UC San Diego
Scientific workflows, Kepler Scientific Workflow System
Big data applications, distributed computing
Reproducible science

Michael Baitaluk, Ph.D.
Research Scientist, SDSC
Scientific data modeling and information integration
Systems and molecular biology, bioinformatics

Chaitan Baru, Ph.D.
SDSC Distinguished Scientist
Director, Center for Large-scale Data Systems Research (CLDS), SDSC
Associate Director, Data Science and Engineering, SDSC
Associate Director, Data Initiatives, SDSC
Data management and analytics, large-scale data systems
Parallel database systems

Hans-Werner Braun, Ph.D.
Research Scientist Emeritus, SDSC
Internet infrastructure, measurement/analysis tools
Wireless and sensor networks
Internet pioneer (PI, NSFNET backbone project)
Multi-disciplinary and multi-intitutional collaborations

Laura Carrington, Ph.D.
Director, Performance, Modeling, and Characterization Lab, SDSC
Application performance analysis & optimization,
HPC benchmarking, workload analysis, modeling via machine learning
HPC power modeling and analysis
Chemical engineering

Sandeep Chandra, M.S.
Executive Director, Sherlock Cloud
Director, Health Cyberinfrastructure Division, SDSC
Compliance (NIST, FISMA, HIPAA)
Scientific data management
Cloud computing
Systems architecture & infrastructure management

Dong Ju Choi, Ph.D.
Senior Computational Scientist, SDSC
HPC software, programming, optimization
Visualization
Database and web programming
Finite element analysis

Amit Chourasia, M.S.
Senior Visualization Scientist, SDSC
Lead, Visualization Group
Principal Investigator, SeedMe 1 & 2
Visualization, computer graphics
Data science, data management, data sharing

KC Claffy, Ph.D.
Director/PI, CAIDA (Center for Applied Internet Data Analysis), SDSC
Adjunct Professor, Computer Science and Engineering, UC San Diego
Internet data collection, analysis, visualization
Internet infrastructure development of tools and analysis
Methodologies for scalable global internet

Yifeng Cui, Ph.D.
Director, Intel Parallel Computing Center, SDSC
Director, High-performance GeoComputing Laboratory, SDSC
Principal Investigator, Southern California Earthquake Center
Senior Computational Scientist, SDSC
Adjunct Professor, San Diego State University
Earthquake simulations, multimedia design and visualization
Parallelization, optimization, and performance evaluation for HPC

Alberto Dainotti, Ph.D.
Assistant Research Scientist, CAIDA (Center for Applied Internet Data Analysis)
Internet measurements
Traffic analysis, network security
Large-scale internet events

Amogh Dhamdhere, Ph.D.
Assistant Research Scientist, CAIDA (Center for Applied Internet Data Analysis)
Internet topology and traffic, internet economics
IPv6 topology and performance
Network monitoring and troubleshooting

Jose M. Duarte, Ph.D.
Scientific Team Lead, RCSB Protein Data Bank
Bioinformatics
Structural biology
Computational biology
Protein Data Bank

Andreas Goetz, Ph.D.
Co-Director, CUDA Teaching Center
Co-Principal Investigator, Intel Parallel Computing Center
Quantum chemistry, molecular dynamics
ADF and AMBER developer
GPU accelerated computing

Madhusudan Gujral, Ph.D.
Bioinformatics & Genomics Lead, SDSC
Processing & analysis of genomics data
Biomarkers discovery
Genomics software tools development
Data management

Amarnath Gupta, Ph.D.
Associate Director, Academic Personnel, SDSC
Director of the Advanced Query Processing Lab, SDSC
Co-principal Investigator, Neuroscience Information Framework (NIF) Project, Calit2
Bioinformatics
Scientific data modeling
Information integration and multimedia databases
Spatiotemporal data management

Amit Majumdar, Ph.D.
Division Director, Data-Enabled Scientific Computing, SDSC
Associate Professor, Dept. of Radiation Medicine and Applied Sciences, UC San Diego
Parallel algorithm development
Parallel/scientific code, parallel I/O analysis and optimization
Science gateways
Computational and data cyberinfrastructure software/services

Mark Miller, Ph.D.
Principal Investigator, Biology, SDSC
Principal Investigator, CIPRES Gateway, SDSC & XSEDE
Principal Investigator, Research, Education and Development Group, SDSC
Structural biology/crystallography
Bioinformatics
Next-generation tools for biology

Dmitry Mishin, Ph.D.
Research Programmer Analyst, Data-Enabled Scientific Computing, SDSC
Research Programmer Analyst, Calit2
HPC systems, virtual and cloud computing
Kubernetes and containers
Data storage, access, and visualization

Dave Nadeau, Ph.D.
Senior Visualization Researcher, SDSC
Data mining, visualization techniques
User interface design, software development
High-dimensionality data sets
Audio synthesis

Viswanath Nandigam, M.S.
Associate Director, Advanced Cyberinfrastructure Development Lab, SDSC
Data distribution platforms
Scientific data management
Science gateways
Distributed ledger technologies

Mai H. Nguyen, Ph.D.
Lead for Data Analytics, SDSC
Machine learning
Big data analytics
Interdisciplinary applications

Michael Norman, Ph.D.
Director, San Diego Supercomputer Center
Distinguished Professor, Physics, UC San Diego
Director, Laboratory for Computational Astrophysics, UC San Diego
Computational astrophysics

Francesco Paesani, Ph.D.
Lead, Laboratory for Theoretical and Computational Chemistry, UC San Diego
Theoretical chemistry
Computational chemistry
Physical chemistry

Philip M. Papadopoulos, Ph.D. (outgoing)
Chief Technology Officer, SDSC
Division Director, Cloud and Cluster Software Development, SDSC
Associate Research Professor (Adjunct), Computer Science, UC San Diego
Rocks HPC cluster tool kit
Virtual and cloud computing
Data-intensive, high-speed networking
Optical networks/OptIPuter
Prism@UCSD

Dmitri Pekurovsky, Ph.D.
Member, Scientific Computing Applications group, SDSC
Optimization of software for scientific applications
Performance evaluation of software for scientific applications
Parallel 3-D fast Fourier transforms
Elementary particle physics (lattice gauge theory)

Wayne Pfeiffer, Ph.D.
Distinguished Scientist, SDSC
Supercomputer performance analysis
Novel computer architectures
Bioinformatics

Andreas Prlić, Ph.D.
Senior Scientist, RCSB Protein Data Bank
Bioinformatics
Structural and computational biology
Protein Data Bank

Peter Rose, Ph.D.
Director, Structural Bioinformatics Laboratory, SDSC
Lead, Bioinformatics and Biomedical Applications, Data Science Hub, SDSC
Structural biology/bioinformatics
Big data applications
Data mining and machine learning
Reproducible science

Robert Sinkovits, Ph.D.
Director, Scientific Computing Applications, SDSC
Director, SDSC Training
Co-Director for Extended Collaborative Support, XSEDE
High-performance computing
Software optimization and parallelization
Structural biology, bioinformatics
Immunology

Subhashini Sivagnanam, M.S.
Senior Computational and Data Science Specialist,
Data-Enabled Scientific Computing Division, SDSC
HPC solutions and applications
Science gateways
Data reproducibility
Computational and data cyberinfrastructure software/services

Shava Smallen, M.S.
Manager, Cloud and Cluster Development
Principal Investigator, Pacific Rim Application and Grid Middleware Assembly (PRAGMA)
Deputy Manager, XSEDE Requirements Analysis and Capability Delivery (RACD) group
Cyberinfrastructure monitoring and testing
Cloud infrastructure tools
Cluster development tools

Britton Smith, Ph.D.
Assistant Research Scientist, SDSC
Computational astrophysics
Software development

Mahidhar Tatineni, Ph.D.
User Support Group Lead, SDSC
Research Programmer Analyst
Optimization and parallelization for HPC systems
Aerospace engineering

Igor Tsigelny, Ph.D.
Research Scientist, SDSC
Research Scientist, Department of Neurosciences, UC San Diego
Computational drug design
Personalized cancer medicine
Gene networks analysis
Molecular modeling/molecular dynamics
Neuroscience

David Valentine, Ph.D.
Research Programmer, Spatial Information Systems Laboratory, SDSC
Spatial and temporal data integration/analysis
Geographic information systems
Spatial management infrastructure
Hydrology

Nancy Wilkins-Diehr, M.S.
Associate Director, SDSC
Co-Principal Director, XSEDE at SDSC
Co-Director for Extended Collaborative Support, XSEDE
Principal Investigator, Science Gateways Community Institute
Science gateways
User services
Aerospace engineering

Frank Würthwein, Ph.D.
Distributed High-Throughput Computing Lead, SDSC
Executive Director, Open Science Grid
Professor of Physics, UCSD
High-capacity data cyberinfrastructure
High-energy particle physics

Ilya Zaslavsky, Ph.D.
Director, Spatial Information Systems Laboratory, SDSC
Spatial and temporal data integration/analysis
Geographic information systems
Visual analytics
Geosciences

Andrea Zonca, Ph.D.
HPC Applications Specialist
Data-intensive computing
Computational astrophysics
Distributed computing with Python
Jupyterhub deployment at scale

SDSC

San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive MC 0505
La Jolla, CA 92093-0505

www.sdsc.edu
twitter/SDSC_UCSD
facebook/SanDiegoSupercomputerCenter