Our National Reach
Partnerships, Communities, and Collaborations

Historically, SDSC has conceived, nurtured, and raised multiple partnerships and collaborations with individuals, communities, and institutions across a wide spectrum of disciplines and fields of study across the nation.

These collaborations have yielded significant published papers and presentations at prestigious scientific meetings which, in turn, have offered new avenues for scientific discovery. To illustrate SDSC's research impact beyond UC San Diego, during the period 2007-2011, the Center received more than $80M in sub-awards from more than 50 non-UC San Diego research institutions, accounting for nearly 90 total awards. Similarly, SDSC allotted nearly $8M in sub-awards to 34 non-UC San Diego research partners during the same time period.

SDSC provides an organizational home for technology and science research and deployment for a wide range of projects to institutions in all regions of the country. Some of these activities include:

- **Partnership in the NSF's Extreme Science and Engineering Discovery Environment (XSEDE).** XSEDE, a partnership of 17 institutions, represents the most advanced, powerful, and robust collection of integrated advanced digital resources and services in the world. SDSC, the only supercomputer center participant on the West Coast, provides advanced user support and expertise for XSEDE researchers across a variety of HPC applications, in addition to support

for the organization's central accounting database. SDSC also offers grid monitoring services through Inca, used by leading grid projects worldwide to identify, analyze, and troubleshoot user-level grid problems and failures.

- **Data-intensive HPC resources.** In 2011, *Trestles* came online to provide researchers from a diverse range of disciplines significant computing capabilities using flash-based memory, which can read data as much as 100 times faster than spinning disk, write data faster, and is more energy-efficient and reliable than standard disk technology. The following year, *Gordon* went into production as the first high-performance supercomputer to use large amounts of flash-based memory, making it the "largest thumbdrive in the world." The result of a five-year, $20 million grant from the NSF, *Gordon* has 300 trillion bytes of flash memory and 64 I/O nodes, making the system ideal for data mining and exploration, where researchers have to churn though tremendous amounts of data just to find a small amount of valuable information. In effect, *Gordon* is designed to do for scientific research what Google does for web searches.
- **OSG Members Gain Access to SDSC Compute Systems.** In June, UC San Diego and the Open Science Grid (OSG), a multi-disciplinary consortium funded by the U.S. Department of Energy and the National Science Foundation (NSF), announced a partnership under which campus researchers gained access to the OSG's fabric of Distributed High-Throughput Computing capabilities. The collaboration was designed to benefit researchers with high-throughput workloads commonly used in domains such as biomedical and life sciences, as well as the geosciences. The partnership means that UC San Diego is one of only a few research universities in the U.S. that is served by both OSG and NSF's XSEDE. It also means that members of the OSG would gain access to two key high-performance compute systems at SDSC, *Trestles* and *Gordon*.
- **Internet research for cybersecurity.** The Cooperative Association for Internet Data Analysis (CAIDA), based at SDSC, began work last year under a contract from the Department of Homeland Security's Science and Technology Directorate (DHS S&T) to help improve security in federal networks and across the Internet, while developing new and enhanced technologies for detecting, preventing, and responding to attacks on the nation's critical information infrastructure. Under this agreement, CAIDA will continue to grow its distributed Archipelago (or Ark) active-measurement infrastructure that currently consists of 62 monitors deployed in 29 countries on six continents. Researchers will use Ark to collect measurements from probes sent to all of the routed IPv4 prefixes on the Internet, and also will experiment with recently developed techniques that improve the efficiency and coverage of IP-level topology probing. The second phase of the project will focus on implementation techniques investigated during the first phase, while the third phase will include demonstration of these new technologies and systems in realistic operational environments, while continuing to add new and replace obsolete Ark monitors.

INNOVATIVE SCIENCE GATEWAYS

Nancy Wilkins-Diehr is an associate director of SDSC and co-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.

Science gateways are used today to provide access to many of the tools used in cutting-edge research—telescopes, seismic shake tables, supercomputers, sky surveys, undersea sensors, and more. A single gateway can give thousands of users access to current, optimized versions of analysis codes at any time. SDSC has been a science gateways pioneer, and has provided innovative projects for XSEDE including the CIPRES Science Gateway (Cyberinfrastructure for Phylogenetic RESearch), representing almost 30 percent of all active XSEDE users (see page 25). During 2013, SDSC received a grant from the National Science Foundation (NSF) to create a software infrastructure for the Neuroscience Gateway. SDSC was also named the lead institute on a NSF planning grant for a Science Gateway Institute that would offer a complete range of services to develop domain-specific, user-friendly, Web-based portals and tools to build science gateways.

"COMET": FOR THE "LONG TAIL OF SCIENCE"

Last October, SDSC was awarded a $12 million grant from the NSF to deploy *Comet*, a new petascale supercomputer designed to transform advanced scientific computing by expanding access and capacity among traditional as well as non-traditional research domains. The new supercomputer, to be deployed in 2015 with another $12 million anticipated during the production phase, will be capable of an overall peak performance of nearly two petaflops (two quadrillion operations per second) and is designed to be part of an emerging cyberinfrastructure for what is called the "long tail of science", which targets a large number of modest-sized computationally based research projects. In effect, *Comet* will be the successor to SDSC's *Trestles* computer cluster, to be retired in 2014 after four years of service. As stated by SDSC Director Mike Norman: "*Comet* is all about high-performance computing for the 99 percent." *Comet* will be a Dell-based cluster based on next-generation Intel Xeon processors. Each node will be equipped with two of those processors, 128 GB (gigabytes) of traditional DRAM, and 320 GB of flash memory. Since *Comet* is designed to optimize capacity for modest-scale jobs, each rack of 72 nodes will have a full bisection InfiniBand FDR Interconnect, with a 4:1 bisection interconnect across the racks. *Comet* also will include some large-memory nodes, each with 1.5 TB of memory, as well as nodes with NVIDIA GPUs (graphics processing units). The GPU and large-memory nodes will target specific applications, such as visualization, molecular dynamics simulations, and de novo genome assembly.

HEALTHCARE INFORMATION TECHNOLOGIES

Dallas Thornton leads SDSC's Health IT division, focusing on delivering scale-out computing, data management, expertise, and support to partners at UC San Diego, other universities, and state and federal sponsors.

SDSC is building a growing Health IT center of excellence around several large grants in the expanding field of data-driven health care. Under an award from the Centers for Medicare and Medicaid Services (CMS) Medicaid Integrity Group (MIG) Data Engine, the Center is developing innovative solutions to detect fraud, waste, and abuse in the U.S. Medicaid programs. In addition, the group supports the National Children's Study San Diego County with secure cyberinfrastructure in its efforts to perform the largest longitudinal study of children and their development ever undertaken. Partnership in these awards extends SDSC's focus on Big Data in an increasingly socially important area. SDSC also is collaborating with David Haussler at UC Santa Cruz to store the Cancer Genomics Hub (CGHub), a large-scale data repository and user portal for the National Cancer Institute's cancer genome research programs, with the goal of targeting anti-cancer drugs to specific tumors and individual patients based on their genetic signatures.