Our Statewide Influence Aligning with UC Principles

"Success in the Information Age can be measured by the precision, power, and breadth of the tools available to an organization, along with the knowledge and the creativity of the people who use them... Combined with the ability to integrate these components into world class 'cyberinfrastructure', SDSC has provided a foundation for research discoveries, education paradigms, and innovative business solutions. SDSC's impact on UC, the State of California, and beyond has been felt in many ways."

Governor Jerry Brown, December 8, 2011

Though SDSC was founded by the National Science Foundation (NSF) almost three decades ago in response to academic researchers nationwide seeking the latest supercomputing resources to solve "grand challenge" problems facing science and society, the Center's mission has evolved over time to include substantial work with, and for, the University of California's initiatives and researchers.

As part of that effort, SDSC has worked to align itself with three principles that "define the goals and purpose that drive and distinguish UC-wide research investments." Those principles are:

- Act as one system of multiple campuses to enhance UC's influence and advantage
- Promote efficient inter-campus collaborations and system-wide economies of scale
- Serve the State of California

Indeed, when SDSC became an organized research unit at UC San Diego in 1996, then Gov. Pete Wilson recognized the Center as a "cornerstone of California's vision for business and academic leadership in science and technology." Since then, SDSC not only has continued its highly successful national mission, the Center also has delivered value and prestige to UC and the State of California through numerous activities with a focus on its resources, services, and expertise in computational and data science. Some notable achievements include:

Richard Moore is SDSC's deputy director, PI of the *Trestles* project, and co-PI of the new *Comet* supercomputer project.

- **Research and discovery.** During the past five years, SDSC staff has collaborated on projects with more than 90 UC researchers across eight campuses. SDSC staff is currently participating in about 50 research grants and proposal collaborations across the UC system.
- **Funds through successful grant applications and research partnerships.** Over its almost 29-year history, SDSC revenues have exceeded $1B—a level of sustained funding matched by few academic research units in the country. SDSC has leveraged the operating funds it receives from UC Office of the President (UCOP) and UC San Diego and has returned six times the amount in sponsored research awards from FY2009 to FY2013.
- **High-performance and data-intensive computing.** Since its launch in 1985, more than 7,500 UC researchers have used SDSC's HPC systems. The Center currently provides compute and HPC storage resources to about 560 UC researchers, including 150 principle investigators across eight UC campuses.
- **Cost efficient colocation and "green" computing.** SDSC has made its 19,000 ft^2 data center available as a recharge-based colocation facility to more than 90 UC groups spanning eight campuses, with an estimated annual system-wide utility cost-savings that exceeds $500K.
- **Innovative cyberinfrastructure.** CI services, available at competitive rates to UC researchers, include the *Triton* computer resource which through 2012 provided more than a million core-hours to more than 600 users across eight UC campuses, and was used for teaching classes both at UC San Diego and UC Santa Barbara; in addition to the SDSC Cloud storage service, among the largest academic cloud storage system in the world.
- **Host to numerous critical databases.** SDSC houses numerous data sources of high impact including: the Protein Data Bank (protein structures); Medicaid data from the Centers for Medicare and Medicaid Services (CMS); the Library of Congress Chronopolis digital preservation system; Open Topography.org (LiDAR data); and the American Red Cross Safe and Well website, developed at SDSC in urgent response to the Hurricane Katrina disaster and the need to match missing people with family and friends. In 2010, SDSC worked with the office of California's CIO to launch the California Spatial Data Infrastructure Project, with 96 Terabyes (TB) of dedicated storage including the California Coastal Atlas which is being used to assess the impact of the sea level rise from climate change along the California coastline.
- **Bridge to empower the next generation.** Through its award-winning TeacherTech program, SDSC has trained more than 1,000 teachers in the San Diego region in science and technology, helping many underserved students to span the "digital divide" to the Information Age.

Aside from these current activities, SDSC has embarked on several initiatives for UC and the State of California, including:

- **Deployment of a massive computing system to support the needs of a larger and more expansive community of scientists and researchers, an activity sometimes referred to as the "long tail of science."** To help fill these needs and continue the Center's local and national focus on high-performance and data-intensive computing, SDSC is building the new *Comet* supercomputer, funded by NSF (see page 42).
- **Establishment of SDSC's brand in data science across UC through the creation of a new institute focused on data science and engineering.** To help fill the need to educate and train a new generation in data science, and continue the Center's innovations in research and discovery, SDSC is developing the Institute for Data Science and Engineering. This institute will develop and provide applied informatics hands-on experience and training across several UC campuses and UC San Diego. Some of the material will be suitable for an Introduction to Data Science undergraduate course offered to UC students system-wide, while other material could be used as part of a graduate curriculum, also across multiple graduate programs.
- **Building research collaborations focused on Big Data solutions for problems of high public interest for the benefit of UC and the State of California.** Under the Center's newly established "Data Initiatives" program, SDSC is reaching out for partnerships and collaborations on both the management and technical aspects of Big Data and other data-enabled applications. Projects under way included the BigData Top100 list, a community-based effort to establish the first global ranking for systems designed for Big Data applications, an effort that came as an offshoot of the Center for Large-scale Data Systems Research (CLDS), formed in 2012. SDSCs colo facility now stores all the cancer genomes from major projects funded by the National Cancer Institute, called CGHub, with ample room for expansion to house and provide secure access to all the genomic data from UC medical centers, and to provide the data analysis that will drive clinical genomic medicine in the future.

ANALYZING WILDFIRE BEHAVIOR

Three research organizations from UC San Diego, including SDSC, were awarded a multi-year National Science Foundation (NSF) grant in 2013 to build an end-to-end cyberinfrastructure to perform real-time-driven assessment, simulation, prediction, and visualization of wildfire behavior.

In addition to SDSC experts, the project—called WIFIRE—includes researchers from the California Institute for Telecommunications and Information Technology (Calit2) Qualcomm Institute, and the Mechanical and Aerospace Engineering (MAE) department with UCSD's Jacobs School of Engineering.

The WIFIRE CI (cyberinfrastructure) was designed to support an integrated system of wildfire analysis, with specific regard to changing urban dynamics and climate. The system integrates networked observations such as heterogeneous satellite data and real-time remote sensor data, with computational techniques in signal processing, visualization, modeling, and data simulations to provide a scalable method to monitor such phenomena as weather patterns that can help predict a wildfire's rate of spread.

The products of WIFIRE will be initially disseminated to project collaborators, including CAL FIRE, the U.S. Forest Service, and SDG&E covering academic, private, and government laboratories while providing value to emergency officials and first-responders, and in turn the general public.

WIFIRE will be available for use by government agencies in the future to save lives and property during wildfire events,

Immersive technologies such as the NexCAVE at Calit2/Qualcomm Institute may play a central role in wildfire incident command center simulation and training. Photo by John Hanacek, Calit2/UC San Diego

test the effectiveness of response and evacuation scenarios before they occur, and assess the effectiveness of high-density sensor networks in improving fire and weather predictions.

"WIFIRE will be scalable to users with different skill levels using specialized web interfaces and user-specified alerts for environmental events broadcasted to receive before, during, and after a wildfire," said Ilkay Altintas, principal investigator for the WIFIRE project.

The WIFIRE CI encompasses the remote sensor network that is part of the High Performance Wireless Research and Education Network (HPWREN) project started at SDSC under NSF funding in 2000. HPWREN director and co-founder Hans-Werner Braun is a co-PI of WIFIRE; in addition to Larry Smarr, founding director of Calit2; and MAE Professor Raymond de Callafon.

"BIGDATA TOP100" TO BENCHMARK "BIG DATA"

Chaitan Baru, an SDSC Distinguished Scientist, was recently named the Center's Associate Director of Data Initiatives. Baru also directs SDSC's Center for Large-scale Data Systems Research (CLDS), focused on technology and technology management issues related to Big Data.

SDSC researchers are coordinating and providing the intellectual leadership toward the creation of a "BigData Top100" list, the first global ranking of its kind of Big Data systems, blending benchmarking approaches from high-performance computing, transaction processing, and database query processing. An initial board of directors has been formed to steer this activity, coordinated by Chaitan Baru, director of the Center for Large-scale Data Systems research (CLDS)—an industry-sponsored center of excellence created within SDSC to develop concepts, frameworks, analytical approaches, and systems solutions to address technical as well as technology management challenges facing information-intensive organizations in the era of Big Data. The board includes representatives from companies in California and beyond including Milind Bhandarkar, Pivotal; Dhruba Borthakur, Facebook; Eyal Gutkind, Mellanox; Jian Li, IBM; Raghunath Nambiar, Cisco; Meikel Poess, Oracle; and Tilman Rabl, University of Toronto.

SDSC ASSISTS RESEARCHERS IN NOVEL WILDLIFE TRACKING PROJECT

Amit Chourasia is a senior visualization scientist at SDSC and Visualization Services Group lead for the Center. Chourasia also is principle investigator for the SEEDME.org project.

A team including researchers from the U.S. Geological Survey (USGS) and the San Diego Zoo's Institute for Conservation Research has developed a novel methodology that for the first time combines 3D and advanced range estimator technologies to provide highly detailed data on the range and movements of terrestrial, aquatic, and avian wildlife species.

Relying on expertise from SDSC researchers, the team created highly detailed data sets and visualizations after they tracked three highly iconic but threatened species in the U.S., southwest China, and northeastern Australia: California condors, giant pandas, and dugongs—a large marine animal somewhat similar to the manatee.

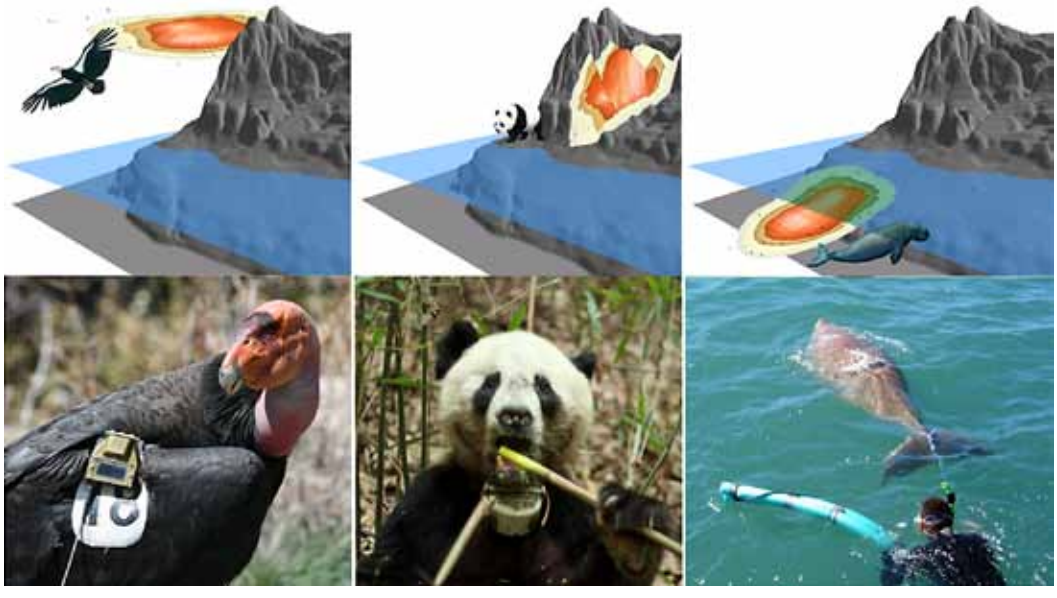
"We were able to speed up their software by several orders of magnitude," said Robert Sinkovits, director of SDSC's Scientific Applications Group, which helps researchers make optimal use of SDSC's larger supercomputers, including *Gordon* and *Trestles*. "In this case, calculations that had formerly taken four days to complete were finished in less than half an hour."

Gordon is being used to host the visualizations, and will make it easier for the software developers to explore the impact of algorithmic modifications on the quality of the solution. "Most importantly, we expect they will use our systems to solve other challenges that were previously considered to be intractable," said Sinkovits.

The visualization expertise was provided by Amit Chourasia, a senior visualization scientist at SDSC. "We made changes to write the data into a more compact format, which enabled swift output and ingestion," said Chourasia. "A key goal was to allow the experts to visualize the data directly on *Gordon* via remote access, as it is essential to minimize data movement and replication when data sizes grow. Currently, we're working to fuse data from various sources such as topography and climate to further aid the understanding of such habitats."

"Our collaborative research team harnessed the power of SDSC to fully exploit the increasing size and quality of 3D animal biotelemetry tracking and datasets," said James Sheppard, an ecologist with the San Diego Zoo's Institute for Conservation Research, and a member of the research team. "This gives us deeper insights into patterns of animal space-use, and informs strategies for the conservation management of endangered species and their habitats."

The study, called 'Movement-based Estimation and Visualization of Space Use in 3D for Wildlife Ecology and Conservation', was submitted to the *PLoS-ONE* online science journal. In addition to Sheppard, researchers for the study included Jeff Tracey, (USGS and lead author); Jun Zhu (University of Wisconsin, Madison); Fuwen Wei (Chinese Academy of Science, Beijing); Ronald Swaisgood (San Diego Institute for Conservation Research); and Robert Fisher (USGS, San Diego).



Example avian, terrestrial, and aquatic animal biotelemetry datasets and their spatial domains and home range contours. Left: California condor with a GPS biologger attached to its patagium. Center: A giant panda telemetered with a GPS collar. Right: A dugong fitted with a tail-mounted GPS biologger.