Fighting COVID-19 with Knowledge Graphs

SDSC’s Peter Rose & Ilya Zaslavsky Receive NSF RAPID Funding

The Office of Integrative Activities at the National Science Foundation (NSF) awarded SDSC Researchers Peter Rose and Ilkay Zaslavsky Funding to organize COVID-19 information into a trans-disciplinary knowledge network that integrates health, pathogen, and environmental data to better track cases to improve analysis and forecasting across the greater San Diego region.

[Continued on page 7]
As I mentioned in our well attended ‘e-General Staff Meeting’ on May 20, the world shifted radically this spring due to the COVID-19 pandemic. We’re likely headed for a new “normal” that will affect many aspects of everyday life for some time. But while so many things have changed, one constant is that as a center, SDSC is still tremendously productive and focused on moving ahead even as we all adapt to new ways of communicating as well as making major adjustments, especially for those with children at home.

Based on weekly situation reports from SDSC division directors, I believe that SDSC continues to fire on all cylinders – or should I say a fully charged battery pack with plenty of range – despite substantial pressure on UC budgets. This issue of Innovators underscores that by highlighting several new grants both related to fighting the pandemic as well as ongoing research enabled by Comet and other resources across all science domains.

In summary, SDSC is well-positioned as an NSF-funded HPC center for the foreseeable future. Our PIs continue to win significant grants, and our recent CloudBank award gives us a strategic advantage going forward as scientific computing evolves to include the public cloud providers.

Selecting recipients for the latest Value Awards announced at the GSM was focused on recognizing those who have gone above and beyond even before the pandemic struck. The Team Award, however, was an obvious choice, recognizing our front line: Keith Green, Tim McNew, Nick Patience, Sophorn Khem, Mark Johnson, Fred Spinney, Anthony Steinell, and Jeff Filliez, for keeping our machine room ops running so we can provide essential support to COVID-19 research on a national scale. Other categories:

- The Leadership Award: To Brian Balderston and Jeff Filliez for guiding essential data center operations during the pandemic.

- The Saved the Day Award: To Sonia Nayak in our Business Office for untangling all the HUBzero contractual complications.

- The Community Building Award: To newcomer Nathan Snodgrass for his role as the only customer interface for the HUBzero community and preserving good customer relationships.

Congratulations to all award recipients, and to all for your outstanding dedication! Please remain safe and well.

Michael L. Norman
SDSC Director
Supercomputers Unlock Reproductive Mysteries of Viruses and Life

Viruses such as the dreaded severe acute respiratory syndrome coronavirus 2 rely on the host cell membrane to drastically bend and eventually let loose the replicated viruses trapped inside the cell. Scientists at Boston University and UC San Diego recently used supercomputers including SDSC’s Comet to develop simulations to help propose a mechanism for this budding off of viruses. In addition, a related study also relied on supercomputer simulations to find a mechanism for how the DNA of all life adds a base to its growing strand during replication. This fundamental research could help lead to new strategies and better technologies to fight infectious and genetic diseases.

The cell membrane remodeling study, important for viral reproduction, cell growth, and other biological processes was published online in February 2020 in the Biophysical Journal. Qiang Cui, a professor in the Departments of Chemistry, Physics, and Biomedical Engineering at Boston University and the study’s co-author, also was part of the study on DNA base addition, published late last year in Proceedings of the National Academy of Sciences. Cui also co-authored a computational study that relied on supercomputer simulations to determine a chemical mechanism for the reaction of nucleotide addition, used in the cell to add nucleotide bases to a growing strand of DNA.

“By doing that, computationally, we are also able to determine the role of a catalytic metal ion of magnesium that’s in the active site of the enzyme DNA polymerase,” said study co-author Daniel Roston, an assistant project scientist at UC San Diego’s Department of Chemistry and Biology. “This metal has been a bit controversial in the literature. Nobody was really sure exactly what it was doing there. We think it’s playing an important catalytic role.”

SDSC’s Comet Supercomputer Helps Benchmark Cancer Immunotherapy Tool

With the American Cancer Society estimating 1.76 million new cases and more than 600,000 deaths during 2019 in the U.S. alone, cancer remains a critical healthcare challenge. In efforts to help mitigate these numbers, researchers at Rice University used SDSC’s Comet supercomputer to evaluate their new molecular docking tool called Docking INCrementally or DINC, which aims to improve immunotherapy outcomes by identifying more effective personalized treatments.

Led by postdoctoral researcher Didier Devaurs, the Rice researchers recently published their evaluation of DINC in the BMC Molecular and Cell Biology journal. The most significant result is that their molecular docking approach can make predictions of molecular interactions that other docking tools would miss. This has strong implications in cases where these predictions are notoriously difficult to make, and especially in the context of immunotherapy, which leverages the immune system to combat cancer.

The goal of their DINC tool is to assist with identifying these peptides for cancer immunotherapy, which required significant high-performance computing resources. "Our study showed that Comet has the computational power to make predictions that could be useful to immunotherapy," said Devaurs. "We will next assess how to rank the numerous predictions it makes to provide only the most realistic ones to our clinician colleagues working on novel cancer treatments."
UC San Diego’s CREATE and SDSC Awarded National K-12 STEM Grant

**Innovation Bloc Grant Part of $75 million DoD Initiative**

The U.S. Department of Defense’s (DoD) Defense STEM Education Consortium has awarded a one-year grant to SDSC and UC San Diego Mathematics Project based at the university’s Center for Research on Educational Equity, Assessment, and Teaching Excellence (CREATE), to introduce computing into high school math classrooms.

The DSEC Innovation Bloc grant, called ICAT through DM – short for Introducing Computing and Technology Through Discrete Math problem solving – will fund a Summer Institute for regional high school mathematics teachers and a Summer Academy for military-connected, underserved, and underrepresented rising seniors.

Robert Sinkovits, director of SDSC’s scientific computing applications and lead for SDSC’s Education and Training program; and Osvaldo Soto, director of the UC San Diego Mathematics Project at CREATE, are the Principal Investigators for the grant. Mathematics is traditionally taught using a pencil-and-paper approach and often lacks a real-world context, frequently leading to a sense of alienation among middle and high school students along with missed opportunities to advance critical mathematical and computational ways of thinking, noted Sinkovits.

“Computational and mathematical thinking are essential to the DoD workforce in nearly every science domain,” said Sinkovits. “Discrete mathematics is particularly important in areas as diverse as cryptography, logistics, communications networks, and complex decision making.”

The DSEC Innovation Bloc grant was awarded as part of a five-year, $75 million award by the DoD under which DSEC focuses on K-16 (Kindergarten through college) STEM enrichment programs for military-connected and/or low-income students and educators, as well as workforce engagement, program evaluation, and public outreach efforts across the nation.

Read more at qrgo.page.link/Nnp6n

**2020 REHS Program Starts June 22!**

The Research Experience for High School students (REHS), part of SDSC’s student outreach initiatives, marks its 11th year next month, as the program gets underway – remotely, like so many other programs and conferences. SDSC Education Manager Ange Mason founded REHS to help increase awareness of computational science and related fields of research to students in the San Diego region. Students are mentored by computational research scientists and other SDSC staff on a wide variety of topics and research projects to gain exposure to career options and work-readiness skills through hands-on experience.

Read more at qrgo.page.link/e9eSj

**SDSC Pilots First Remote GPU Hackathon**

The first completely remotely attended GPU Hackathon with SDSC successfully concluded on May 13, marking a new chapter in the evolution of NVIDIA’s hackathon program. By early March the planning group, including representatives from National Energy Research Scientific Computing (NERSC), Oak Ridge National Laboratory (ORNL), SDSC, and other institutes discussed the possibility of doing the event remotely instead of onsite at SDSC, despite myriad challenges of doing so given the large number of participants from multiple time zones.

In early March, within weeks of accepting applicants, the decision was made to explore the possibility. “It was SDSC’s turn to host the next hackathon, and things were changing daily due to COVID-19,” said Susan Rathbun, SDSC’s Program and Events Manager. “Given the number of GPU hackathons on the calendar, rescheduling one for later this year at SDSC was not a very viable option.”

SDSC Staff Research Associate Cindy Wong, who works with Susan, did a deep-dive on Zoom to determine what was feasible, and that’s when things started to gel. To help explain the new format and familiarize participants with an entire suite of tools and resources, the planning team introduced a pre-event training session two weeks before the event.

“The training event made a huge difference in terms of getting people up to speed,” said Mary Thomas, a computational data scientist and HPC training lead at SDSC, who participated as a co-organizer and mentor, adding that the overview tutorials helped overcome any concerns.

Read more at qrgo.page.link/uAypa
In April, SDSC lost two alumni that many of you may remember fondly. Ann Redelfs, who joined SDSC in 1995 as the Center’s director of external relations before returning to Cornell University in 2001, passed away on April 18 in Duluth, MN. Nancy Jensen, part of SDSC’s Industrial Relations team before joining the Center’s External Relations group where she worked with Ann Redelfs, passed away in Desert Springs (Tucson, AZ) on April 26, 2020.

Ann was a major force in the evolution of SDSC,” said SDSC Founding Director Sid Karin. “She was instrumental in establishing the organizational culture at the heart of much of our success. Ann also played a similar leadership role in the broader high-performance computing community, as well as in the larger computing community.”

Jennifer Matthews, who was part of Ann’s External Relations and EOT team, called her a “profoundly decent and generous human who found a way to lift up everyone around her. With a gracious smile and a metal-toed boot, Ann threw her full self into efforts that empowered women and minorities in science, and was the consummate mentor. Her passing reaches far beyond the HPC community.”

Redelfs’ tenure with NSF’s supercomputer centers program, always in external relations, began in 1986 at the Cornell Theory Center, one of nation’s first four supercomputer centers in addition to SDSC. From there, she directed external relations for the Center for Research on Parallel Computation at Rice University in Houston, TX, before joining SDSC. She was named director of education, outreach, and training (EOT) in 1999.

A Facebook page has been set up for the ‘Spirit of Ann Redelfs (SOAR)’. Planting a tree in memory of Ann can also be made at the Arbor Day Foundation’s Trees for Others program.

Nancy retired from General Atomics in 2002 and then finished her career at SDSC in 2008. She worked with Ann Redelfs in External Relations and was involved in many of the major proposals that supported the evolution of the Center. Our own Sandy Davey shared some wonderful memories of one of her best friends and colleagues:

“Nancy was born in Massachusetts but as a young child moved to San Diego with her mother. She would share fond memories of cow pastures in Mission Valley as she saw firsthand the growth of San Diego. She never drove on the freeway, yet she knew how to get anywhere in San Diego County by using the back roads. She loved to explore every inch of this beautiful city.

“Nancy coordinated numerous special events, which led her to her role as a certified Conference Planner. She was the most organized person I ever met. With her planner in hand, she had a way of organizing that just put one at ease – plus she had the most beautiful penmanship I’ve ever seen!

“Nancy was SDSC’s first Team Captain for “Super Striders” for ACS Making Strides Against Breast Cancer. She was an inspiration for many of her fellow workers to raise significant dollars for ACS for many years to come. She understood the purpose of networking and was great at bringing people together.

To make a donation in Nancy’s memory, please visit www.pmdalliance.org/give/
In August 2019 UC San Diego, UC Berkeley, and the University of Washington were awarded a five-year, $5 million grant by the National Science Foundation (NSF) to develop CloudBank, a suite of managed services to simplify public cloud access for computer science research and education. The award marked the first NSF-funded public-private partnership for cloud use as a pilot project for future collaborations.

CloudBank will help the NSF by bundling multiple small requests that come directly to them into a bulk request to cloud providers, thereby reducing the more costly direct connections. Through this aggregation and innovative financial contract type, CloudBank will pass along savings to researchers that would otherwise be unavailable to them.

Shava Smallen, a research programmer/analyst with SDSC’s Data Enabled Scientific Computing (DESC) group, was named CloudBank co-PI, with SDSC Director Michael Norman as the program’s PI.

Q: You’re asked to describe CloudBank’s goal in two sentences. What would they be?
The CloudBank portal essentially combines cloud account management, cost monitoring, outreach, and training. The program’s goal is to make it easier for computer science researchers and educators to use public cloud resources.

Q: Are there any restrictions on who can use CloudBank resources and services?
Allocations on public clouds will only be provided to projects that belong to NSF programs that have specified the use of CloudBank. So far there have been two such solicitations, and we expect to see awardees from these solicitations come into CloudBank in the coming months.

Q: How it actually works is probably more complex given all the moving parts.
CloudBank, with our partner Strategic Blue, is handling all of the minutiae of how money moves around between UC San Diego and cloud providers on behalf of PIs. Once their access to clouds is set up, a research team can use the CloudBank portal to view their funds, track and monitor their use of cloud services, optimize their usage, and ultimately ensure the most efficient use of their cloud funds. The financial operations of CloudBank have their roots in UC San Diego’s ITS Division. Under the leadership of Vince Kellen, UC San Diego’s CIO and one of the CloudBank Co-PIs, the campus has been innovating ways how universities and researchers work with public cloud providers.

Q: How is CloudBank addressing what are called certain ‘pain points’ for researchers?
Several of the pain points that we identified during the proposal focused on the amount of time it takes to get access to clouds, learning about the available cloud services and how to get started, and managing costs. So, CloudBank will be providing (1) automated accounting/billing setup and configuration to facilitate the onboarding process, (2) a cloud services catalog and curated training materials to help researchers get started with cloud resources, and (3) a multi-cloud cost monitoring tool (Nutanix Beam) and automated email notifications to facilitate cost management. Researchers can also log into the CloudBank portal and cloud consoles using their institution’s credentials due to our integration with CILogon.
Q: You're about two-thirds of the way through what would be called Year One of the program. Can you tell us what's been done to-date and how the CloudBank User Portal is progressing?

We have started early testing with a handful of researchers and are looking to expand that this month. CILogon integration was recently completed for each of the initial set of cloud providers: Amazon Web Services (AWS), Google Compute Platform (GCP), and Microsoft Azure. The accounting/billing setup and configuration is ready for AWS since our campus had a lot of experience with them coming into the project. GCP and Azure are still in progress, but should be ready by the end of June thanks to assistance from Google and Microsoft. There’s also an initial version of the cloud services catalog that’s being reviewed by the cloud providers, which will also help us maintain and keep the entries up to date. An initial set of training videos that introduce new users to cloud concepts and technologies is also under development, and we’re working on curated training content for each of the providers.

Q: What happens during the next four years?

Once we go into production this August 1, CloudBank will provide ongoing support for researchers to help them prepare cloud cost estimates for their NSF proposals, start using cloud services for their research, learn how to run classes in the cloud, and troubleshoot issues. CloudBank will also be holding annual workshops and presenting information and training at related conferences. Several portal development efforts will continue as we learn and respond to researcher needs, provide more fine-grained cost monitoring notifications, and refine procedures for offboarding researchers toward the end of their awards.

Q: This sounds like a tremendous undertaking. When not immersed in the clouds, what do you do to get yourself grounded again?

I’m very grateful that my life has only been minimally impacted by the pandemic so far. Thinking about the many others who are not as fortunate is very grounding. Also, like others in the SDSC community, I love cycling and this has been a great way to get outside lately, especially with fewer cars on the road. I’ve been exploring new routes and getting to know my neighborhood better. Connecting virtually with friends and family has also been essential during this time.

For more about CloudBank, please see www.cloudbank.org or email help@cloudbank.org

Fighting COVID-19 with Knowledge Graphs (Continued from page 1)

The award, funded under the NSF’s Rapid Response Research or RAPID program, began in May and will run through the end of October. It will allow the researchers to quickly launch a comprehensive semantic integration platform for data-driven analysis and development of policy interventions that take into account up-to-date health, social-economic, and demographic characteristics of populations in different areas; biomedical information such as virus strains and genetic profiles.

This project will be coordinated with another RAPID program led by Krzysztof Janowicz, a professor of Geographic Information Science at UC Santa Barbara, with focuses on infrastructure resilience, supply chain disruptions, and local policy decisions designed to combat the pandemic.

“The project will be based on our knowledge graph prototype linking information about pathogens, health data, and environmental indicators and enabling cross-domain inferencing,” said Rose, director of SDSC’s Structural Bioinformatics Laboratory and principal investigator (PI) for the project, called ‘COVID-19-Net: Integrating Health, Pathogen and Environmental Data into a Knowledge Graph for Case Tracking, Analysis, and Forecasting.’ “Such a graph lets researchers trace the spread of the coronavirus in different geographic conditions, focusing on specific virus strains and transmissions.”

“The main goal is to make these datasets easier to find, index, and integrate,” said Zaslavsky, director of Spatial Information Systems Laboratory at SDSC and UC San Diego, and co-PI of the new award. Zaslavsky also specializes in geoscience data discovery and develops visual exploratory gateways into advanced data science and machine learning tools, which will serve as one of the user dashboards for querying the graph.

Specifically, the project will refine the knowledge graph and integrate it with other complementary graphs being developed by the Open Knowledge Network (OKN) that was funded in September 2019 under the NSF’s Convergence Accelerator program. Tasks include refining methodology for populating the knowledge graph using Jupyter-based data-to-knowledge ingestion pipeline, continuously extending the knowledge graph content with additional data, and developing efficient data markup using recent schema.org COVID-19 extensions.

Read more at qrgo.page.link/Yu9SM
Technology Forum Breakfaists Go Virtual

Readers of this column may be familiar with the SDSC Technology Forum Breakfast series, which brings together our industry partners, staff, researchers, and friends to start their morning with breakfast and a presentation showcasing a technology or project of interest. Through these events, we value the in-person interaction and ability to build relationships and discuss new ideas and potential collaborations. However, with no clear view in sight as to when we may be able to resume in-person meetings, we made the decision to go virtual with these for the foreseeable future, sans breakfast, unfortunately.

We held our first virtual event on May 28, with Frank Würthwein, Professor of Physics at UC San Diego; and Igor Sfiligoi, Lead Scientific Software Developer at SDSC, discussing their experiment to spool up a “virtual, multi-cloud GPU supercomputer.”

Over a weekend last fall, using the Open Science Grid (OSG) platform, Würthwein and his team acquired all available GPU computing instances worldwide from three major commercial cloud providers to assemble a coordinated computing environment that conducted actual scientific analysis on data from the IceCube Neutrino Observatory, a large-scale scientific instrument occupying a cubic kilometer of natural ice in Antarctica. For a brief period of time, the experiment controlled about 50,000 GPUs and achieved 380 petaflops (32-bit precision) of performance, rivalling that of some of the world’s top supercomputers on high-throughput computing (HTC) workloads.

In this column’s view, this work has important implications for both science and industry regarding Infrastructure as a Service (IaaS) for large-scale computing. So until we meet again as the saying goes, we hope to see you on Zoom July 16 for our next ‘virtual breakfast’ meeting, which will feature AMD speaking about their latest processor technologies for high-performance computing.

For information and registration, visit industry.sdsc.edu/events

Ron Hawkins
SDSC Director of Industry Relations