June 19, 2018 | By Warren Froelich

West Big Data Hub at SDSC to Partner for Data Storage Network under New NSF Grant

Project Will Help Researchers Share Data More Efficiently, Enable Collaborations

The West Big Data Innovation Hub (WBDIH) at the San Diego Supercomputer Center (SDSC) at UC San Diego is one of four regional big data hubs partner sites awarded a $1.8 million grant from the National Science Foundation (NSF) for the initial development of a data storage network during the next two years. Other partners include Johns Hopkins University and University of Chicago, awarded a $300K EAGER for Open Storage Network (OSN) software.

The team will combine its expertise, facilities, and research challenges to develop the OSN. The demonstration project will result in the design of a larger, low-cost, scalable national system capable of being replicated across many universities. The OSN will enable national collaborations and allow academic researchers across the nation to share their data more efficiently than ever before, according to the NSF announcement.

"We are excited to support OSN to help meet the needs of researchers in today's era of data-driven discovery and innovation," said Erwin Gianchandani, acting assistant director of the NSF's Computer and Information Science and Engineering Directorate. "The OSN team and their supporting collaborators will build a community to multiply the impact of previous and current NSF investments and anchor comprehensive data infrastructure that will be vital to the future of our nation's scientific and engineering enterprise."

The project, led by Alex Szalay of Johns Hopkins University, leverages key data storage partners throughout the U.S. including the National Data Service and members representing each of the other three NSF-funded Big Data Regional Innovation Hubs: the Midwest Hub at the National Center for Supercomputer Applications (NCSA) at the University of Illinois Urbana Champaign, the South Hub at the Renaissance Computing Institute (RENCI) at the University of North Carolina, and the Northeast Hub at the Massachusetts Green High Performance Computing Center and the Pittsburgh Supercomputing Center (PSC).

"The OSN is bringing together the nation's advanced and supercomputing centers to create a new, distributed platform that will provide several services for researchers including for data sharing and cloud caching," said Christine Kirkpatrick, executive director of the National Data Service, WBDIH deputy director, and division director of IT Systems and Services at SDSC. "The Big Data Hubs are a perfect coordinating and outreach point for ensuring the technical solutions are met with smart approaches to governance and sociotechnical challenges."

NSF's investment in OSN builds on a seed grant by Schmidt Futures – a philanthropic initiative founded by former Google Chairman Eric Schmidt – to enable the data transfer systems for the new network. These systems are designed to be low-cost, high-throughput, large-capacity, and capable of matching the speed of a 100-gigabit network connection with only a small number of nodes. This configuration will help to ensure that OSN can eventually be deployed in many universities across the U.S. to leverage prior investments and establish sustainable management for the overall storage network.

"We are excited to support Professor Szalay's promising work designing and testing these impressive storage devices, and want many such open-design petabyte units to be assembled and deployed in and for universities," said Stuart Feldman, chief scientist at Schmidt Futures. "We applaud NSF's investment in the Open Storage Network as a key step toward enabling research requiring truly massive amounts of data."

OSN builds on NSF's longstanding leadership and investments in data science. The new storage network aligns with one of the 10 Big Ideas for Future NSF Investments Harnessing the Data Revolution (HDR). HDR offers a profound opportunity to use data to transform and advance discovery and innovation across all fields of science and engineering. In particular, OSN will help address the HDR goals to build data infrastructure for research and advance fundamental data-centric research and data-driven domain discoveries.

The user experience is an important component of OSN. The new storage network will be piloted by researchers at participating institutions to ensure that it is easy to use, has adequate performance, can be efficiently accessed from various parts of the internet, employs good security and privacy policies, is highly reliable, and has a long duration for data preservation. Additional software and service layers will be added to OSN as it is developed. The NSF is also funding a project led by Ian Foster at the University of Chicago to explore the use of Globus services, which are already widely used for data management with OSN.

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit https://ucsdnews.ucsd.edu/media-resources/faculty-experts.