

From Data to Discovery to A Better World: The National Consortium for Data Science (NCDS)

CONTACT DETAILS:

NAME: STAN AHALT

ORGANISATION: RENCI (UNC-CHAPEL HILL)

PROJECT NAME: NATIONAL CONSORTIUM FOR
DATA SCIENCE

EMAIL: AHALT[AT]RENCI.ORG

Objectives

Launched in early 2013, the National Consortium for Data Science (NCDS) is a U.S.-based public-private partnership that seeks to secure the leadership role of U.S. business and academia in data science research and innovation. It's vision is to provide the structure and leadership that will cultivate a focused data science community that can solve big data problems, drive the field forward and unleash the power of data to enable economic growth. To achieve its vision, the consortium has identified five key goals: 1) Engage a broad community of data science experts to identify key data science challenges; 2) Coordinate data science research priorities and encourage research that spans industry, academia, and government as well as multiple disciplines; 3) Facilitate the development of data science education programs that train a new generation of data experts; 4) Support the development of technical, ethical, and policy standards for data; and 5) Apply NCDS expertise and resources to data challenges in clinical genomics, social and behavioral population studies, water and climate sciences, material sciences, and the industrial Internet. As the consortium grows and matures, its programs, expertise, and partnerships will serve as the foundation for a National Center for Data Science.

On-going activities:

Although the consortium is less than six months old, it has already launched programs designed to forward its vision. The annual invite-only NCDS Leadership Summit brings together data and domain scientists to discuss strategies for translating data into practical innovations and competitive advantages. The first Leadership Summit, held in April 2013, focused on the data challenges in genomics and produced a strategic white paper. A second summit is planned for Spring 2014. The NCDS is launching a Data Observatory, which brings together compute, storage, and data management tools and makes them available as hands-on teaching tools at academic member institutions. A pilot project this fall includes several courses at member universities. The observatory will lay the foundation for a Data Laboratory, which will provide data science researchers access to emerging tools and the physical infrastructure to test new techniques for storing, sharing, analyzing, transforming, and visualizing data. A Data Fellows program will support and mentor data researchers at member campuses and encourage them to collaborate with NCDS members in government, industry and academia. In addition, the consortium plans to host lectures and outreach events aimed at reaching STEM students and connecting them to NCDS members.

Results:

The National Consortium for Data Science is only in its infancy, but already is producing results. The Leadership Summit, *Data to Discovery: Genomes to Health*, brought together the scientists who create and use data (genomic scientists in this case) with the technical experts who build and deploy data management, analysis and storage infrastructure. The Summit white paper was produced in mere weeks and is currently under review by a major research journal. The long-term plan is to publish a summary version of the paper in a journal or technical publication and publish the full paper on the NCDS website. The process will be repeated each year with new summits

focused on data challenges in other domains and new white papers that will drive the development of data science policy and new data science research. The Data Observatory effort is already producing at least two courses that will give students the chance to work with real-world large, heterogeneous data sets. The Data Fellows program, expected to launch this fall, will support talented data science researchers. Their research results will advance data science, build the scholarship of the NCDS, and further its long-range goals and vision.

URL: www.data2discovery.org