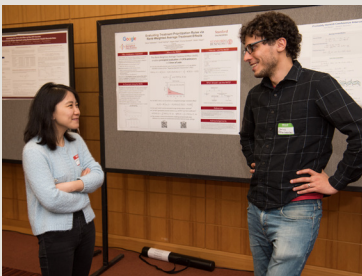


Stanford Data Science Director's Report

FALL 2022–23



Cultivating a campus-wide community of data science leaders

World-class faculty

In fall 2021, Stanford Data Science (SDS) joined the entire Stanford community in celebrating [Guido Imbens](#) (Economics, Graduate School of Business), faculty director of SDS Causal Science Center (SC²), for [receiving the Nobel Prize](#) in economics. Both Guido and [Matthew Gentzkow](#) (Economics), a fellow SC² executive committee member, were later elected to the National Academy of Sciences.

Although faculty recruitment efforts in 2021 did not yield a hire, we are leveraging insights from the process in three searches underway now—with the School of Humanities & Sciences, the School of Engineering, and Wu Tsai Neurosciences Institute. We expect that our joint hires will benefit Stanford broadly as interdisciplinary searches become increasingly common.

Talented, early-career scientists

In fall 2022, we welcomed [14 new Data Science Scholars](#) bringing the [current cohort](#) to 31 graduate students. Graduate students from all seven Stanford schools are eligible to apply; SDS awards the selected students half of their fellowship support for two years. Scholars form a community of practice where they can share and learn from one another while advancing data science solutions for their unique domain challenges. The newest Scholars work in classics, health policy, and computer science, among many other disciplines.

Ongoing strong demand for the program and the success of its alumni are among the measures of its impact. Over the past two years, SDS has received an average of more than 100 applicants for 13 open positions. Since 2018, SDS has supported 70 Scholars, including 39 alumni who went on to academic postdoctoral and faculty positions, industry jobs, and founding commercial start-ups.



Guido Imbens with Graduate School of Education Visiting PhD candidate Kristoffer Ibsen at the student-led Causal Science conference.

We also welcomed [three new SDS Postdoctoral Fellows](#) whose investigations include climate and health research, the evolution of urban growth, spatial inequality, and housing vulnerability in the past fifty years. Fellows typically receive support for two years during which they are mentored by Stanford faculty and build new data science methods that further their domain research. In fall 2022, SDS received more than 70 applications for three open postdoc positions. Of the 12 finalists, one-third do research that addresses population health /sustainability.

For both programs, SDS targets graduate students and postdocs who resist traditional disciplinary boundaries and can make meaningful contributions through the cross-pollination of ideas. Our robust community of future data science professionals and leaders bolsters Stanford's strengths in domains like sustainability, computer science, statistics, and the life sciences, ensuring these fields can meet the challenges of the twenty-first century.



SDS Scholar [Thomas Teisberg](#), BS '18, PhD '25 (Electrical Engineering), is addressing the lack of data that hinders our ability to predict how the melting Antarctic ice sheet will affect sea level changes. As a member of the [Radio Glaciology Lab](#) in the Stanford Doerr School of Sustainability, Thomas helps develop robotics with ice-penetrating radar to expand subsurface data collection. His research relies on data-driven approaches and physics-informed machine learning to assess the ice sheet's behavior and optimize resource deployment.



We still don't understand how brain-gut interactions, including those of the autonomic nervous system, contribute to disorders like Parkinson's, diabetes, or chronic migraines. [Sandya Subramanian](#), SDS Postdoctoral Fellow (Bioengineering) and 2022 [Schmidt Science Fellow](#), unravels these connections through a blend of computer science, statistics, and physiology. She seeks to improve our ability to monitor and quantify these physiologic processes in nonclinical environments, enabling clinicians to personalize therapeutic and management strategies.



Can data science leverage the wisdom of crowds to make scientific expertise—in topics like climate science or public health—more credible to lay audiences? When could this backfire? SDS Scholar [Shilaan Alzahawi](#), PhD '24 (Organizational Behavior, Graduate School of Business), is bringing the fields of data science and organizational behavior together to understand if reforms like open or “crowd science” improve scientific reproducibility and credibility. As an active SDS member, Shilaan is an affiliate at the Center for Open and Reproducible Science and former advisor for Data Science for Social Good.

Building and empowering a research ecosystem

Centers and communities of practice

SDS now has a network of four centers to accelerate data science across campus, including the newly launched [Sustainability in Data Science](#) (SuDS), led by Stanford Doerr School of Sustainability faculty member [David Lobell](#) (Earth System Science). SuDS aims to create engaging interactions to inform data scientists of the big questions in sustainability and ensure sustainability scientists across campus are aware of the latest capabilities of data science. In addition to David, center activities are led by faculty in the Stanford Doerr School of Sustainability, the School of Engineering, and other scholars from diverse fields. SuDS launched with a cosponsored seminar series on [wildfire research](#).

In fall 2021, the [Center for Open and REproducible Science](#) (CORES) hosted a lecture series featuring Stanford faculty, SDS Scholar Shilaan Alzahawi, and [Armin Thomas](#), a Ram and Vijay Shriram Data Science Fellow. It also hosted a [spring lecture series](#) showcasing how to use technical tools like Python and R in the research process. In March 2022, CORES held its second annual conference ([full video here](#)) where it presented awards to both faculty and trainees to acknowledge their contributions to open science.

In early 2022, CORES Faculty Director and SDS Associate Director Russ Poldrack began engaging with HELIOS, a group of universities collaborating to promote a more transparent, inclusive, and trustworthy research ecosystem within and across their campuses. Under Russ' leadership and building on SDS's support of CORES, President Marc Tessier-Lavigne and Provost Persis Drell, along with Vice Provost and Dean of Research Kathryn "Kam" Moler, committed resources to appoint a director of open scholarship, a newly created position based in CORES.

The Causal Science Center hosted its [inaugural student-led campus conference](#) in fall 2021, featuring presentations from 16 graduate students and a keynote by faculty director Guido Imbens. The center continues to cosponsor its long-running weekly [Online Causal Inference Seminars](#), with several recordings available [here](#).

Led by James Zou (Biomedical Data Science, Computer Science), the [Data4Health](#) center seeks to create an inclusive and university-wide community of researchers and students using data science to extract biomedical and healthcare insights. James aims to increase collaborations as well as provide shared datasets and computational infrastructure for biomedical researchers.

Note: The [SDS Collaboratory](#), an early model for SDS centers, wrapped up its grant from the National Science Foundation and prepared a final report on outcomes, including efforts to foster collaborative research teams and brainstorming working groups with faculty, students, and postdocs.

Enabling data-intensive computation and discovery

Over the past three years, SDS has recognized that faculty whose research leverages machine learning and artificial intelligence would benefit from frictionless access to data-intensive computational power—and the expert personnel needed to optimize it. The SDS Advisory Group encouraged SDS to champion a resource that would provide faculty with the hardware, software, and—critically—staff research data scientists needed to enable data-driven discovery at scale.



As envisioned, this computational and data science instrument would enable new discoveries across disciplines spanning health, physics, and materials (among others), and in doing so help advance university priorities like the new Stanford Doerr School of Sustainability and the Stanford Institute for Human-Centered Artificial Intelligence. As

such, it would complement existing resources such as Sherlock (Stanford’s shared computing cluster), national labs, and cloud services, which would continue to play important roles.

In October 2022, SDS wrapped up a faculty task force and submitted a report to key Stanford leaders including President Marc Tessier-Lavigne, Provost Persis Drell, and Vice Provost and Dean of Research Kam Moler. Informed by interviews with more than four dozen faculty and university leaders, along with experts in high-performance computing, the report laid out the potential cost savings entailed by this new on-site resource, how it would strengthen Stanford’s ability to attract and retain top faculty and students, and its critical role in educating students and scholars for future positions in academia and industry. The report will now be considered in the context of a broader assessment of campus computation needs led by Kam Moler.



WiDS cofounder and faculty emerita Margot Gerritsen interviews Susan Wojcicki, CEO of YouTube, and Diane Tang, MS '98, PhD '01 (Computer Science), Google Fellow, at the 2022 WiDS Conference.

Expanding our reach

Convening a broad community

SDS held its inaugural conference in April 2022, which featured opening remarks from President Marc Tessier-Lavigne and a keynote on data-intensive research on the neuroscience of pain and surgical pain management by Emery Brown, professor of anesthesia, Harvard Medical School and Massachusetts General Hospital, and professor of medical engineering and computational neuroscience, Massachusetts Institute of Technology. Other highlights include the introduction of SDS faculty-led centers and a panel discussion on improving trust in data science with [Debra Satz](#), Dean of the School of Humanities & Sciences. More than 230 community members joined in-person and 2,000 virtually. A conference video is available [here](#).

SDS was pleased to welcome the Women in Data Science (WiDS) initiative to its community in September 2021, when it moved from its founding home in the School of Engineering. In March 2022, the program hosted its seventh annual world-wide conference with over 400 in-person at Stanford and 10,000 virtual attendees. The on-campus event was complemented by 200 regional events in 55 countries, which attracted 27,000 in-person and 130,000 virtual attendees. A conference recap including videos of sessions can be found [here](#). SDS hired a new director of engagement and partnerships who will further integrate WiDS into the SDS ecosystem.



Emery Brown

WiDS SPONSORS

Boeing
General Motors Corporation
IBM
Intuit
Meta
Microsoft Corporation
Pinterest
SAP
Systematica Investments LP
The Mathworks Inc.
TotalEnergies
Walmart

Training next-generation data scientists

SDS Associate Director [Chiara Sabatti](#) (Biomedical Data Science) offered BIODS260 in Spring 2022 for the second time, instructing Stanford students on how to mentor students from underrepresented groups interested in learning data science and its applications to biology, medicine, and healthcare. [Balasubramanian “Naras” Narasimhan](#), SDS senior research data scientist, and [John Chambers](#), SDS senior advisor, taught a [new graduate course](#) on computing for data science, featuring speakers from around the country and topics ranging from privacy preservation to evaluating the accuracy of machine learning models.

In fall 2022, Stanford introduced a new undergraduate major aimed at enhancing data literacy across all disciplines. Chiara Sabatti co-led the creation of a new Data Science Bachelor of Science (which replaced the Math and Computational Science major). Students can also earn a Bachelor of Arts in Data Science & Social Systems. Both tracks of the data science major boast a robust core curriculum ranging from statistical inference to ethics, and require a capstone project where students are expected to apply what they have learned. More details can be found in this [Stanford News article](#).

Over the summer, six undergraduate and master’s students from Stanford and beyond participated in the [2022 Data Science for Social Good](#) (DSSG) program. The cohort tackled two projects: one accounting for undocumented COVID-19 deaths, and another tracking police activity and accountability in incidents involving substance abuse and mental health crises. Under the guidance of Stanford researchers and technical mentors, participants in the eight-week program worked closely with governments and nonprofits as they learned how to apply data science to real-world social problems. This was the fourth year of DSSG at Stanford. Since its launch, more than 32 student fellows (undergraduate and master’s students) and 16 technical mentors (PhD students and postdocs) have participated in the program.

Our future home

In 2022, Stanford broke ground on the Data Science and Computation Complex. Located adjacent to the Quad and the Oval, the building is intended to foster interdisciplinary computation and data research through a flexible framework of permanent offices, rotating research team spaces, and collaboration areas—all designed to adapt and evolve with these rapidly evolving fields. SDS looks forward to joining other units in this modern space, including the Department of Statistics, the new undergraduate major in Data Science, the Brown Institute, the Department of Computer Science, and the Symbolic Systems program.

In June 2022, SDS Postdoctoral Scholar alumnus Dallas Card received a distinguished paper award at the American Computing Machinery conference on Fairness, Accountability, and Transparency (FACCT) for “The Values Encoded in Machine Learning Research.” His coauthors include two researchers in Stanford’s Department of Computer Science. Dallas is currently an assistant professor in the School of Information at the University of Michigan.

“The most special aspect of the Data Science Scholars program is the community. It is unusual to be with students and postdocs with such diverse academic backgrounds. I have not only learned from listening to the presentations of other students, but learned to communicate my research to people outside of my field. Thanks to SDS, I felt empowered to pursue opportunities like co-organizing and teaching a workshop on ‘Data management for publishing’ for graduate students and Stanford, as well as teaching a session in the Science Coding Immersion Program at San Francisco State University.”

—**Fatima Alejandra Pardo Avila**, *Structural biologist, Data Science Scholar alumna*

“SDS fills a critical gap. Through SDS, researchers from divergent fields can convene and see first-hand how others solve similar problems—and that’s one of the most effective ways to accelerate progress. Moreover, making science more rigorous and transparent is essential to maintaining public trust in research. But knowledge about doing that varies widely even at Stanford. Through CORES, I am expanding access to tools and expertise so that all researchers can maximize their effectiveness and impact.”

—**Russ Poldrack**, *Alfred Ray Lang Professor of Psychology, SDS Associate Director, Faculty Director of SDS Center for Open and Reproducible Science (CORES)*

“SDS provided me with opportunities to explore research areas that are related to but different from my PhD research—which prepared me well for my future interdisciplinary academic career. It also gave me the opportunity to mentor/collaborate with PhD students from both statistics and environmental engineering fields.”

—**Xiao Wu**, *Assistant Professor of Biostatistics at Columbia University, SDS Postdoctoral Fellow alumnus*