

## 2020 ADSA Data Science Leadership Summit Session 1 Update from the University of California, Berkeley

### Our route toward establishing organizational units for data science

Berkeley's strength in the diverse areas of data science has been growing for decades, and that growth has accelerated along with the field in recent years. See our [timeline of progress](#) preceding the announcement of the Division of Computing, Data Science and Society (CDSS).

### Our vision for the path ahead

*Our job is to weave together the riches of the university to solve societal problems. We are focusing on climate and sustainability, biomedicine and health, and social welfare and social justice. We're not here just to advance our core capabilities (though we will do this too), but to integrate expertise from around the campus to advance new research agendas and to ensure that students come out of Berkeley thinking masterfully and ethically about data so they can transform whatever field they choose to enter.*

— Jennifer Chayes  
Associate Provost, CDSS, September 10, 2020

### Milestones reached since the last ADSA DSL Summit

#### Institutional progress

- Announced [CDSS](#), which includes the Data Science Education Program, the School of Information, and the Berkeley Institute for Data Science (BIDS), and involves the Departments of Statistics and Electrical Engineering and Computer Sciences (EECS). It also will feature the Data Science Commons, a new and dynamic organizational structure for faculty who are advancing transdisciplinary discovery. (Feb)
- Received [\\$252 million gift to seed construction of a new “data hub”](#) – the largest gift in UC Berkeley's history – where the Division will be housed. (Feb)
- Completed **appointments of key executive staff** for the Division: Associate Dean of Faculty, Associate Dean of Graduate Education, Associate Dean of Undergraduate Education, Chief Administrative Officer, and Chief Development Officer. (Feb-Aug)

#### Undergraduate and graduate education

- Drew over 500 colleagues from across the nation to 3<sup>rd</sup> annual [National Data Science Education Workshop](#), sharing best practices and creating a community for instructors of data science education. (Jun)
- Launched [Data6](#), a new six-week summer intensive precursor to Data 8. (Jun)
- Graduated
  - **331 undergraduate students with a degree in Data Science**, of which 143 were double majors, and **40% were women**.
  - **168 total graduates in data science master's programs**, as well as 29 students completing the new [Graduate Certificate in Applied Data Science](#) through the School of Information.
  - **240 undergraduates majors, 52 masters students, and 15 doctoral students** from the Department of Statistics, including [this year's winner of the University Medal](#), UC Berkeley's highest honor for a graduating senior

- 730 from EECS, which has been selected to host the [Rising Stars 2020 Academic Careers Workshop for Women](#) (May)
- “Graduated” **nine BIDS Data Science Fellows** (PhDs) who advanced to academic and research positions with UC Berkeley, Smith College, the Halicioğlu Data Science Institute, and UC San Diego, as well as with startup companies including Anomaly and Ambys Medicines.

### **Research awards, fellowships, and partnerships**

- Awarded
  - [\\$12.5 million NSF grant with MIT](#) to establish a **multidisciplinary institute to improve understanding of critical issues in data science**, including modeling, statistical inference, computational efficiency, and societal impacts.
  - [\\$10 million NSF grant](#) to gain a theoretical understanding of **deep learning**.  
These are both led by faculty from the Departments of Statistics and EECS. (Aug)
- Awarded five-year NIH grant for **computational social sciences training program**, which is led out of BIDS and D-Lab. Welcomed first cohort of five graduate student fellows from epidemiology, public policy, social welfare, and sociology into the program. (May)
- Announced our [second multi-year strategic industry relationship with Accenture Applied Intelligence](#) and BIDS to support interdisciplinary research and training for exploring major social and scientific challenges, such as ethical AI, biomedicine, and environmental sustainability in California. Relationship includes **new postdoctoral data science fellowships** in environmental sciences and social justice. (Mar)
- Awarded an NSF grant to **evaluate and expand our data science curriculum to ensure it is accessible** to a broad range of students. D-Lab oversees the project, [Undergraduate Data Science at Scale](#) project, which includes Mills College and the University of Maryland, Baltimore County (UMBC). (Nov ‘19)
- Welcomed inaugural cohort of Data Science Health Innovation Fellows, as part our [first multi-year strategic industry relationship with Johnson & Johnson Innovation](#) and BIDS. (Oct ‘19)

### **Rapid response data science - research and outreach efforts**

Faculty, data science researchers and fellows, and research software engineers across CDSS have undertaken research projects in response to the pandemic and Black Lives Matter global movement. As early as April 15, we curated a **COVID-19 data repository** for forecasting county-level death counts in the United States. As recent as September 15, we developed an ambitious proposal for a California **law enforcement accountability network** to address police misconduct.

To reach well beyond CDSS, we also inaugurated the [COVID-19: Berkeley Conversations - Computing and Data Science in Action series](#), addressing topics such as how to blunt the coronavirus, support a robust recovery, and not unintentionally rely on data and algorithms to guide pandemic response that may actually serve to perpetuate these inequities.

### **What’s been and remains challenging**

#### **About the “intrinsic interdisciplinarity of the field”**

1. Planning the [Data Commons](#) - an experimental space designed to advance new interdisciplinary programs that may not fit into the footprint of existing departments.
2. Evolving our interdisciplinary research engagement models to best serve the changing and multiplying expectations of faculty and data science researchers.

#### **About being (still) a start-up in increasingly challenging economic and social times**

1. ‘Building the bench’ to sustain data science capacity for rapid response.
2. Cultivating new donor relationships to achieve extremely ambitious fundraising goals.
3. Establishing a shared identity among the data science academic and research units.