Harvard Data Science Review • Issue 3.2, Spring 2021

Data Science and Computing: The View From a Sister Campus

Hal S. Stern¹ Debra J. Richardson² Marios Papaefthymiou³

¹Department of Statistics, Donald Bren School of Information and Computer Sciences, University of California Irvine, Irvine, California, United States of America,

²Department of Informatics, Donald Bren School of Information and Computer Sciences, University of California Irvine, Irvine, California, United States of America,

³Department of Computer Science, Donald Bren School of Information and Computer Sciences, University of California Irvine, Irvine, California, United States of America

Published on: Jun 30, 2021

DOI: <https://doi.org/10.1162/99608f92.1a7d19ab>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

The perspective of a top public university on a topic, data science and computing, that is central to higher education in the 21st century is extremely valuable. We are grateful to Jennifer Chayes for her well-articulated discussion of the centrality and impact that data science and computing are having on society (“[Data Science and Computing at UC Berkeley](#),” this issue). The explosive growth that is occurring in this area makes it a critical topic for universities around the world to address. The achievements at Berkeley are most impressive; we can expect other campuses to benefit from the work done there, although not all will choose to or be able to follow the Berkeley model. There is no doubt in our minds that a variety of approaches are useful in addressing society’s needs for continuing innovation in data gathering, analysis, and interpretation strategies, as well as strategies for developing a data-aware workforce. *Harvard Data Science Review* provides a great service by inviting additional perspectives, and we are pleased to provide a report of our experiences.

At UC Irvine (UCI), a sister campus to UC Berkeley, campus-wide efforts on data science and computing are centered in the Donald Bren School of Information and Computer Sciences (ICS). ICS includes three departments: Computer Science, Informatics, and Statistics, and a number of research centers. We write as the three individuals that have served as dean of ICS since it became a school in 2002; each of us represents a different one of the three constituent departments. The historical development of ICS is informative about the data science and computing perspective of our campus. The (then) Department of ICS was formed in 1968; as computer science was developing as a field, there was a recognition at Irvine of the importance of the human components associated with computing and thus a desire to provide a degree of separation from the School of Engineering. The department developed expertise in traditional computing disciplines like theory and algorithms, databases, graphics, and software engineering. It also developed expertise in human–computer interaction and the social and organizational impacts of computing, and welcomed interdisciplinary collaborations that were unusual at the time. Statistics developed late at UC Irvine; the department was founded in 2002, at roughly the same time that the burgeoning Department of ICS was transitioning to a school with multiple departments. It was natural to include statistics, a discipline with both information and computational elements, in the new school.

The combination of the three key elements, computing, statistics, and information, in a single academic entity, as Berkeley is building, provides a number of advantages. At Irvine, we share a common physical environment (a building houses the large majority of our faculty) and work together often on hiring committees, seminar series, grant proposals, and so on. Collaborations in teaching, research, and program development are easily managed in such an environment. We briefly describe the experience at UC Irvine in building up educational programs, carrying out impactful research, and partnering with our colleagues on campus.

Educational Programs

Having the computer science, informatics, and statistics departments in a single school makes developing educational programs straightforward. University resources often depend on student enrollments within a

school. Partitioning credit for programs that are joint with other academic units can be a challenge or impediment, though one that can usually be worked out. Not having the challenge enabled us to set up a data science major relatively quickly; our program began admitting students in 2016. Our B.S. in Data Science program is focused on developing students with strong skills in mathematics, statistics, and computing who are well equipped to be partners in developing novel methods to address a range of data types and to participate in collaborative teams using data to inform science and policy. Upon graduation, the students have the background to assume positions in industry or pursue graduate school in either statistics or computer science if they have that desire. The program includes course sequences in statistical methods and calculus-based statistical theory. It also includes a range of core computer science topics like algorithms, data structures, data management, and computer organization. The program concludes with a senior-level requirement for a two-quarter (20-week) team-based project course. This capstone requirement reinforces the critical role that teams play in successful efforts to address science and policy questions.

To get beyond our major and reach out to other students on campus, UCI has developed its own Introduction to Data Science course. With no prerequisite, the course introduces the full data cycle. Topics include data collection and retrieval, data cleaning, exploratory analysis and visualization, introduction to statistical modeling and inference, and communicating findings. Applications include real data from a wide-range of fields following reproducible practices. Students learn to make predictions and inferences using models and consider the impacts of data-based decisions.

More recently, UC Irvine has developed a master of data science program (MDS) aimed at developing a cadre of well-trained data professionals to support the needs of industry and government. This full-time program can be completed in approximately 15 months. With minimal prerequisites (programming, calculus, linear algebra, and introductory probability/statistics), the program provides background sufficient to allow students to identify data relevant to a scientific or policy question, apply the appropriate statistical and computational skills to gather the data, and implement relevant analytical procedures over large, heterogeneous data sets in a modern cloud computing environment. Like the undergraduate program, a key element is a capstone course that provides additional training in writing and communication. This new program is welcoming its first cohort in the Fall of 2021.

At the doctoral level, the Ph.D. programs in all three departments allow students to focus on different aspects of the data and computing world. Our students easily enroll in courses across the departments to develop the skills they need to address their research questions.

Research

ICS at UC Irvine has a long history of leadership in data-centric research issues, machine learning, and artificial intelligence. For example, from its earliest days in the late 1960s, ICS offered graduate coursework in artificial intelligence. This was quite rare at the time. In 1987, ICS graduate students set up the [UC Irvine](#)

[Machine Learning Repository](#) containing data sets that the research community can use to benchmark new methods or to support educational efforts. The archive has been cited more than 1,000 times. Recently refreshed through funding from the National Science Foundation, the repository contains over 500 data sets.

Faculty in ICS carry out highly regarded research programs in the core disciplines central to data science and computing. Rather than describe these in detail, we focus here on the collaborative research partnerships that complement core research activities in statistics, machine learning, artificial intelligence, and informatics. These partnerships involve other researchers at UCI, researchers at other universities, and corporate partners. There are partnerships across the range of disciplines that Chayes describes in her Section 5. Notable initiatives include the Connected Learning Lab, the Center for Statistics and Applications in Forensic Evidence (CSAFE), and the Institute for Genomics and Bioinformatics. The Connected Learning Lab includes faculty from six different academic schools at UCI studying and mobilizing learning technologies in equitable, innovative, and learner-centered ways. CSAFE is a research partnership of a number of universities focused on developing novel methods to accurately assess forensic evidence and thereby assure a more equitable justice system. The Institute for Genomics and Bioinformatics has been around for nearly 15 years fostering innovative basic and applied research at the intersection of the life and computational sciences. Its members partner with UCI researchers in medicine and biological sciences on issues associated with personalizing medical treatments.

The centers listed are just a few examples. Faculty in ICS cover the campus with their partnerships and collaborative grants and projects. They are central members of nearly all of the extramurally funded interdisciplinary collaborations on campus that are the hallmark of modern science. The latest innovation in data science and computing at UC Irvine is the philanthropically funded Irvine Initiative in AI, Law, and Society, which brings together faculty, graduate students, and researchers from across campus (especially in the UCI Schools of Law, Social Science, and ICS) focused on the promise and challenges of the increasing use of algorithms in science and policy.

Challenges and Opportunities

The academic world is a place of great opportunity for those with interests in data science and computing and their combined impact on science and policy. As described by Chayes and reiterated here, there are countless examples where large, heterogeneous, and diverse data sets are changing the way research is done in the physical sciences, the life sciences, the social sciences, the humanities, and so on. As described above, modern research teams require collaboration of data science and computing researchers with those in a wide range of disciplines. In many areas, data sets are also posing novel questions about society—How can we ensure that data and algorithms are inclusive and equitable? How can we ensure that data are secure? How can we protect the privacy of data sources? Answering such questions also requires partnerships of individuals from many disciplines. Yet, all too often, university communities evaluate their members by the standards of individual disciplines. We need to make sure that faculty reviews and promotions appropriately value the interdisciplinary collaborations that are the key to scientific breakthroughs.

A challenge that we have found at UC Irvine is that a limiting factor for advancing data science collaborations on our campus—and we imagine this is true at many universities—is that there are not enough people with the data skills and computing expertise to meet the needs of campus research teams. At UC Irvine this challenge is partly mitigated by our Center for Statistical Consulting. More broadly, however, the demand for well-trained data professionals makes educational programs like the ones at UC Berkeley and the ones at UC Irvine essential for society. The demand is there now; we should not delay in building programs that create a cadre of well-trained students with expertise in data science and computing.

Disclosure Statement

Hal S. Stern, Debra J. Richardson, and Marios Papaefthymiou have no financial or non-financial disclosures to share for this article.

©2021 Hal S. Stern, Debra J. Richardson, and Marios Papaefthymiou. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.