Executive Summary

New data collection technologies and computational and inferential approaches to data analysis are transforming research, decision-making, and action. In the world around us, the capacity of data science to shape human life stretches from precisely targeted medical treatments to integrated reconfigurations of urban experiences to global forecasts of climate change. Within the university, the availability of a vast array of forms, scales, and sources of data has impacts in nearly every academic domain. Building on Berkeley's decades of investment in the foundational and applied areas relevant to data science and our academic excellence across the board, the Faculty Advisory Board of the Data Science Planning Initiative believes our university has a platform to pioneer a comprehensive strategy for the field unmatched by any other institution. As researchers and educators situated across the disciplines, we see a groundswell of engagement among our colleagues and students, and we view this as the moment to act.

Putting a Berkeley stamp on data science means tackling its challenges across the depth, breadth, and richness of the field. In depth, the Berkeley vision for data science must mark our university as the intellectual center from which innovations continue to emerge across foundational areas such as data-oriented systems, algorithms, artificial intelligence, statistical machine learning, and statistical inference, addressing wide-ranging questions that reach from causal inference to graph theory to optimization and controls. In breadth, it entails advancing data science methodologies that are developed through engagements with many disciplinary areas, underwriting progress from engineering across the sciences and social sciences to the humanities and ramifying into professional education. In richness, it obligates us to provide leadership in examining human interactions with data, from deep interpretative questions about the experience of the individual user to the widest ethical and societal challenges inherent in the pervasive deployment of data. The Berkeley model for data science can build on our innovative undergraduate curriculum, which has been recognized widely as a model, and the broad platform available in the Berkeley Institute for Data Science. It should expand to include Berkeley's signature strategy for integrated excellence along the axis linking undergraduate and graduate education with research and extending into high-quality professional education. It must leverage the powerful philanthropic and revenue-generating possibilities around our industry impact in data analytics, inference, and computing, while maintaining our foundational values of openness, diversity, and public mission.

The Faculty Advisory Board sees a clear path to deepen Berkeley's strengths, connect them across campus, and secure global leadership. We urge the university to undertake a rapid set of interdependent measures.

1. **Organizational form: Move to create a flexible, innovative academic core of independent decanal stature**, a School centered on computation and data science, **with a mandate to develop a robust culture of engagement and strong mechanisms of connection campus-wide**. After considering multiple options, we

recommend that the goal be to form this School administratively as a Division of two existing colleges (Engineering and Letters & Science), although it could also exist as a free-standing academic unit. Departments, programs, and institutes can use regular faculty governance processes to populate it; our strong sense is that a world-leading school can be built.

2. **Strategic academic plan for faculty FTE: Invest in an influx of data science faculty positions** (over 10 years, on the order of 20-25 FTE strength). Faculty positions are key to both expanding core domains and building broad strength as this area surges. **Provide a path for faculty to identify targeted application areas for decisive investment of FTE**. Immediate opportunities are in the social sciences in their intersection with computation and data science. We see the need for a next-generation strategy for data-intensive biology and significant possibilities around data science and environment; there will be other emerging areas as well.
3. **Fundraising: Move data science rapidly forward as a central pillar of fundraising** across Berkeley, including new endowed faculty FTEs, support for key programs and institutes, and the new School as a whole.

For this effort to succeed, all three measures are necessary and need to be taken together. To maintain the momentum we now have, they should begin without delay. In advance of creating the academic core unit and executing a search for a dean, we urge the formation of a Data Science Initiative as a transitional vehicle to operate for roughly 24 months, under the leadership of a faculty director tasked to partner with the faculty, the administration, and the Academic Senate to move forward on each of the measures above.

In our era of innovative instruments, pervasive computation, vast new streams and sources of data, and increasing interdigitation of research domains, the impact of Berkeley's pioneering approach to data science will be profound. Our university is poised to build the world's strongest, best-integrated set of programs for data science, allowing us to tackle challenges in research and education in a fashion that does justice to their depth, breadth, and richness. Only in this way can we attract and retain stellar faculty who can form the basis of Berkeley's continued excellence and secure our leadership across this field for many years to come.

Data Science Planning Initiative
Faculty Advisory Board
Report

University of California, Berkeley
2016

Report outline

Part I provides the FAB's overall approach to our task from vision through recommendations. It provides conceptual and narrative orientation leading up to our recommended next steps.

> Part I lays out the vision that we think can draw Berkeley together in data science (Section 2). It reflects on lessons to be drawn from our recent experiments in campus-crossing organizational forms (Section 3). It articulates four expectations that should be placed on any initiative to be taken by Berkeley in data science (Section 4) and summarizes the FAB's recommendations and next steps (Section 5).

Part II lays out the rationales for our recommendations. It offers models for implementation, discusses particular opportunities and challenges, and suggests processes that can follow. **The key elements of the thinking that led to the FAB's recommendations are documented here.** It operates at a greater level of detail than Part I, but it is not supplementary material.

> Part II addresses in more detail the organizational rationales around a strong core in the shape of a new School (Section 6) with an ethos of community and openness and strong relationships connecting it outward (Section 7). It analyzes the need for investments of faculty FTE overall (Section 8) and identifies possible strategic foci, whether these are located inside or outside or across the boundary of the new school (Section 9). It draws out concepts for data science as a fundraising priority (Section 10) and identifies revenue generation possibilities that a new core unit would enable (Section 11).

Part III addresses the institutional context for our recommendations.

> Part III addresses situational challenges that Berkeley would need to address in moving ahead as a campus (Section 12). It outlines a schematic set of processes that would need to follow for our goals to be achieved (Section 13). It closes with concluding remarks about the opportunities open to us (Section 14) and several appendices providing additional details.

Table of Contents

PART III

APPENDICES

1. Introduction

Living in a connected world, where most aspects of individual and societal activity have a digital aspect, where researchers can easily integrate diverse sources across culture and time, where they can gain unprecedented visibility through sensors and imaging and can seamlessly harness vast computational power and analytical methods, and where human expression can instantly engage the entire world, we are immersed in data and its analyses as never before. New capacities to collect, link, and analyze vast datasets have given us deeper understanding of complex systems from gene networks to ecosystems to transportation infrastructures. They shed light on matters whose scale and subtlety range from socioeconomic effects at the margin of perceptibility to far-reaching consequences of climate change. We have enabled new levels of visibility and granular understanding of brain function, urban planning, and medical decision-making and opened up powerful avenues of intervention in individual human behavior and political campaigns. In working with vast noisy data, our ability to extract significance has let us discover supernovae exploding far off in the universe and identify public health interventions that can make a difference in a child's life. Data-driven modeling has enabled computers to mirror human intelligence and underwritten new industries whose business models depend on analyzing the data we each supply. Our data have also been used to track and target us in ways we experience along a spectrum from empowering to disturbing, depending on how we stand in relationship to the organizations in control of the information. Underneath each of these instances is the capacity to manipulate and understand streams of data using technologies and methods that would have boggled the mind even a decade ago. The effects that follow from the ubiquity of data and the power of analytics are being felt throughout research, business, government, society, health, law, and individual lives.

Harnessing computation, connectivity, and analytics offers insight. It also raises deep questions about coming shifts in our work and our world. Berkeley has already powerfully shaped these developments, providing intellectual and practical leadership in key areas as data science has emerged as a national priority and a broadly recognized force in the world. In recent years, the pervasive availability of data has made itself felt inside our university in skyrocketing demand from our students and transformations in our research across a wide swath of terrain. As we match up the signals within the university with indicators outside, we believe that these changes call for a concerted and visionary response. We think it is urgent to build from our existing ground-up excellence to connect across Berkeley and expand our capacity to lead in this domain.

We see a path to do this in a way that brings benefits to the campus by strengthening our faculty and investing in organizational change. As presented in detail below, our report lays out our analysis, our recommendations, and the rationales leading to them, based on our work as the Faculty Advisory Board of the Data Science Planning Initiative. We have taken the charge given to the FAB—to develop an integrated strategy for Berkeley's global leadership in data science—with a seriousness that comes out of our experience as researchers and educators across many disciplines. We sense that this is a moment like Berkeley's reorganization of biology in the 1980s, of similar significance to the campus, for attracting and retaining world-leading faculty and enabling them to do transformative work. This moment matters for Berkeley; it also matters beyond it. This is because when data science is pursued collaboratively and comprehensively in a manner worthy of the world's best public university, it can shape the way the rest of the world engages in this transdisciplinary field.

Data science in the world

The transitions we are seeing around us go beyond incremental change—more data as just more of the same. Across fields, the character of data flows has shifted with massive efforts of digital data generation, particularly since the rise of the Internet and the pervasive connectivity of sensors, devices, and instrumentation. In the last decades we have been transitioning from relying mainly on modes of acquisition of data that were highly selective, structured, and human-directed to mobilizing new forms that can be near-continuous, pervasive, unfiltered, tacit, and variable. These flows can now be worked into domain-area models with real-time, adaptive updating and multi-scale demands on fidelity. Important leaps forward have come from systematic efforts undertaken by teams of researchers to make a multiplicity of data sources available to be fused and integrated. Data extends well beyond the classical format of collections of numerical values. It takes on speech, images, and other formats that, even when provided in digital modalities, have semantic properties and structures of their own. With new streams of unruly, ill-structured, and context-bound data come challenges about provenance, integration, and standards. With new computationally enabled approaches come shifts in underlying conceptual frameworks, ways of engaging with uncertainty, and paradigms of sense-making and learning in the analytical realm. With all of these come deep questions about the human entanglements of this work, given the unprecedented explosion of detailed data on human beings as they live and move through their world.

The term "data science" is a shorthand for the intellectual and practical challenges of bringing together all these modalities of data, their computational handling, and their analytical manipulation to underwrite inferential conclusions and actions in a "datafied" world. The phrase has been used for at least a decade in industry, and there is enough clarity about its meaning for estimates of the impact of data science on the economy to have been made (three trillion dollars annually, globally, from open data alone) and for data science to have become a meaningful career path across a range of industries (early-stage predictions suggested an estimated half a million positions available by 2018, more than half of them unfilled).[1] Major foundations and national funding agencies have marked out data science as a centerpiece of programmatic efforts, with new cross-directorate leadership roles created at NIH and NSF. Four years ago, the White House launched a significant government-wide effort, the first of the multi-

[1] "Open data: Unlocking innovation and performance with liquid information," McKinsey & Company, 2013; "Big data: The next frontier for innovation, competition, and productivity," McKinsey & Company, 2011.

agency Presidential initiatives, to develop data science technologies, demonstrate their applications, and train the next generation of data scientists. President Obama recruited the first Chief Data Scientist to the White House, made open data the new default for federal agencies, and earlier this spring released his administration's Big Data Research and Development Strategic Plan.

Data science in universities

Data science has crystallized around real-world challenges with data, but as an intellectual field it is not limited to them. It involves principled and reliable deployment of abstract reasoning in a world of data that sits at the edge of our ability to handle it, data that is particular, messy, context-bound, and concrete. Rather than just a set of localized practices, data science has deep intellectual foundations, and it calls on and inculcates deliberate habits of thought and principled approaches to technical challenges. It harbors intellectual complexities, moreover, that will take decades of work. These can be found today within foundational areas such as computing and statistics, two fields whose mutual embrace has provided many of the regulative underpinnings on which data science has been based; along research frontiers reaching out into areas of application, where concrete problems are generative of new methods and knowledge; and in disciplines that take on its human complexities and societal embeddedness, bringing in approaches that link data science back to other domains.

It is fair to recognize the institutional challenge presented by data science as an example of interdisciplinary or integrative challenges more broadly.[2] The transdisciplinary character of data science is profoundly anchored in the nature of its work, and this character needs to be consciously accommodated for it to flourish in an academic setting. That said, data science is not just the next interdisciplinary field looking for a foothold. Because of its methodological character, it is more like a common platform than the base of a hierarchy upward. It is different from nanoscience or neuroscience, for instance, because it is foundational in a different way, methodologically rather than materially. Its closer analogues are mathematics or language. And yet the particularities of working with data entangle it unmistakably in contexts of origin and application, no less than in human and societal issues. That is true even as its principles, abstractions, and people are mobile.[3]

---

[2] This report is not invested in the word "interdisciplinary"; for many purposes, "transdisciplinary" works as well. What matters is that the term points to processes that can be at odds with discipline-based university structures. Research universities are institutional creations of the industrial age of the late nineteenth and early twentieth centuries, when specialization, division of labor, and discipline became the social model of knowledge work as well. But the world does not naturally exist in the disciplines the university cuts it up into; it intrinsically exists as a whole.

[3] A 2015 workshop on data science education concluded, "It is clear that that both in research and in teaching, new opportunities created by access to large amounts of data also required bringing together multiple specialties in ways that did not exist before. The change needed is more substantial than simply identifying multiple disciplinary players and dividing responsibilities between and among them. We need to reorganize and reconfigure the disciplines themselves in a way that matches the new reality."

Whatever we make of it, the impact of data science, in the world and intellectually, has begun to be recognized in academia, with different terminology and emphases across the most diverse of domains. It is spreading broadly and moving quickly, without one single model on view about how it should be supported and grown. This process has been underway at other universities over roughly the same timeframe as at ours. The outcome is that there are dozens of programs, institutes, and schools now in place, and new initiatives are getting launched at significant scale. These aim to tackle research challenges in data science, to provide education and training, and to anchor data science in the faculty body, often with significant philanthropic and government commitment in place. These efforts have been moving fast, and there is a great diversity of approaches among them, ranging from hub-like interdisciplinary institutes to institutional change at the scale of creating new colleges. There is clearly no one "proven" model for how to institutionally support and grow data science. And yet experimentation is moving apace because computation and data form a growth area that manifestly matters to the future strength of universities at large.[4]

Arguably, however, there is yet to emerge an academic response at a scale and scope that is commensurate with the transformational nature of data science for society, technology, research, and individual lives. In this free space lies an opportunity: to develop a plan for investing in data science in a way that serves the whole institution at the same time as its constituent parts.

Berkeley's moment: why us, why now

Berkeley was early to invest in data science with a prescient set of faculty hirings along the axis from statistics to computing. We have achieved both world-leading academic excellence and a long-standing track record in industry that are major assets for us now. We are situated in the San Francisco Bay Area, where data analytics has flourished like few other places in the world, in close proximity to Lawrence Berkeley National Laboratory and its advanced capacity in data science and technology. Further, our university has strength across many disciplines in which large-scale data collection is prominent. It has the intellectual reach across the humanities and social sciences that allows it to critically examine the ways in which data science shapes societies and lives. In recent years it has created multiple masters programs that provide professional training, including a degree in the School of Information that is targeted directly at aspiring data scientists. In its internal dynamics, Berkeley has seen a growing set of efforts to shape responses to strong student interest inside current doctoral programs and to create data-intensive platforms across department boundaries in the life, environmental, and social sciences.

Most strikingly, Berkeley faculty have been engaged in a kind of facilitated bottom-up experimentation toward integrative campus-scale programs. In the research domain, we have seen faculty come together across the university to develop the Berkeley Institute for Data

[4] Some of us on the FAB have experience with or lines of sight into other universities' initiatives for data science. These insights, together with informal reviews of the peer landscape we have undertaken, have informed our sense of the institutional options.

Science, with several hundred researchers engaged along the stages of a visioning process that excited the foundations that ultimately invested in BIDS. Over the last two years, we have collaboratively created a model for a comprehensive undergraduate data science curriculum with a core-and-connections structure that can serve as a platform on which other academic programs can build. We have seen the curriculum move from blue-sky vision to start-up at scale with significant appreciation among our students and peers.

Those experiences have illuminated the depth of interest and the felt need for data science, as well as the anchoring of both of these in dynamics that are not entirely served by the academic structures we now have. They have brought faculty into campus-wide collaboration and underlined that our greatest strength is in connection, if we can pull it off. They have underscored the power of Berkeley's existing advantages, when we get rapid institutional alignment behind the direction that our faculty are going. Finally, they have brought home to us that we stand at a moment when things can go either way. If we take seriously Berkeley's industrial and philanthropic connections, both locally and more broadly, and our opportunities for generating revenue to support our operations, we can greatly strengthen the core of this field and enrich the campus at large. If we do not invest, we can expect to accelerate losses of core faculty and rapidly erode away a base that has taken years of ground-up effort to build.

Multi-decade trajectories for institutional success are set at moments like this. In the evolution of academic disciplines, the closest analogue lies in the options exercised by universities half a century ago for the then-nascent field of computer science. Institutions that invested strategically in computer science at that moment secured their future trajectories. Even as they were obliged to design for flexibility in a breathtakingly dynamic domain, the programs they built have persisted strongly in character and strength. On the other side of the coin, in Berkeley's own terms the closest analogue is the reorganization of biology in the 1980s, our last instance of university-level structural change. Refactoring the departments in the biological sciences responded to the emergence of new intellectual needs and the growing incongruity of old institutional forms. Its long-range consequences, through the creation of QB3 and the recruitment of the next generation of world-leading faculty, have been critical to the strength of those disciplines and of Berkeley at large.

In relation to data and computing, we are at a comparable moment. That means we must take action and strategically shape the future we want to help bring into being. At the same time we must build in flexibility, experimentation, and differentiation. The effort is not without risk. That risk must be strategically managed, and it must be placed against the very real risk we incur by choosing not to act.

Scope, structure, and aim of this report

This report represents the reflections of fifteen faculty across the university who agreed to serve on the FAB in 2015-16. We did our work with full consciousness of the advisory responsibility that the campus invested in us, serving not as representatives of departments and programs but as faculty who were asked to help articulate a shared vision that could serve the campus at large.[5] Whether as participants or observers of data science, we have seen the groundswell of engagement in this area. In our review process we have collected information and spoken to colleagues and students. What we have heard and concluded is documented here.

Our report is structured in three parts.

Part I provides the FAB's overall approach to our task from vision through recommendations. It provides conceptual and narrative orientation leading up to our recommended next steps.

> Part I lays out the vision that we think can draw Berkeley together in data science (Section 2). It reflects on lessons to be drawn from our recent experiments in campus-crossing organizational forms (Section 3). It articulates four expectations that should be placed on any initiative to be taken by Berkeley in data science (Section 4) and summarizes the FAB's recommendations and next steps (Section 5).

Part II lays out the rationales for our recommendations. It offers models for implementation, discusses particular opportunities and challenges, and suggests processes that can follow. **The key elements of the thinking that led to the FAB's recommendations are documented here.** It operates at a greater level of detail than Part I, but it is not supplementary material.

> Part II addresses in more detail the organizational rationales around a strong core in the shape of a new School (Section 6) with an ethos of community and openness and strong relationships connecting it outward (Section 7). It analyzes the need for investments of faculty FTE overall (Section 8) and identifies possible strategic foci, whether these are located inside or outside or across the boundary of the new school (Section 9). It draws out concepts for data science as a fundraising priority (Section 10) and identifies revenue generation possibilities that a new core unit would enable (Section 11).

Part III addresses the institutional context for our recommendations.

> Part III addresses situational challenges that Berkeley would need to address in moving ahead as a campus (Section 12). It outlines a schematic set of processes that would need to follow for our goals to be achieved (Section 13). It closes with concluding remarks about the opportunities open to us (Section 14) and several appendices providing additional details.

It must be made clear that with the FAB's current composition, it **cannot** once and for all lay out the composition and vision of a new School. That definitional work is in the hands of the faculty who must come together to constitute it. Likewise, the FAB **cannot** prescribe in detail what a new School's educational offerings, fundraising strategy, revenue generation plans, and cost

---

[5] The FAB's composition is outlined in Appendix A1. The Data Science Planning Initiative of which the FAB is a part was constituted by the administration in Summer 2015 as a response to the ground-up efforts in data science research and teaching across Berkeley emerging over the previous years.

structure must be. The FAB can scope out possibilities and help shape them conceptually, but responsibility for projections and decisions is placed in the hands of faculty who must steward the next stage. Finally, the FAB **cannot** lay out how every other part of the university that uses data science should leverage the pieces of this plan to build out its own strength. Identifying the FTE strategy and philanthropic alignment that can work best for a college, school, division, or department is the province of faculty and deans, while developing the most rewarding forms of relationship between the new School and other academic units must be done case-by-case as works best for the two sides.

Our report lays out the rationales for a set of paths of development, and it does preparatory work assessing what the costs and benefits would be. It aims to set up a process that can follow once the Faculty Advisory Board's work is done.[6] Some of the work of the Data Science Planning Initiative has already begun to have effects, and we will be delighted if the FAB has been able to contribute to moving those next steps ahead. We are glad that it has been possible explore this terrain collaboratively with development staff located in UDAR (University Development and Alumni Relations) and in some academic units on campus. With the DSPI we hope to have helped define an inclusive campus community in data science and created stronger pathways for connection of units across campus through seeding new conversations, building cooperative relationships among programs, and shepherding the first stages of the Data Science Education Program.

[6] As a faculty body, the FAB was charged to focus on matters relating to the academic mission of the university. It was not asked to advise on other important questions the campus will need to address in relation to data-intensive research and teaching campus-wide. Thus our report does not address what infrastructural, IT, or other investments the campus is now making or may need to consider, nor does it directly bring in strategic planning now underway in the University Library in relation to data services. We urge that these areas be integrated into the next stages of discussion.

Part I

2. The Berkeley Vision for Data Science

We see the potential for university-spanning excellence in data science, realized in a way that does justice to Berkeley's character and strengths. Our vision for data science at Berkeley encompasses all of its dimensions: laying the foundations of the field and pushing its conceptual frontiers (we call this **deep**), applying established or emerging technologies and techniques to the wide range of areas or domains (**broad**), and studying the implications of the explosion of data and analysis for ethics, policy, society, and human knowledge (**rich**).

Berkeley's unique chance lies in further strengthening each of these aspects. This section outlines the likely opportunities this comprehensive approach would afford. Further, because of the trajectory that has brought the campus to this point, we have the momentum to do so in a way that builds the constituent elements of data science powerfully *and* keeps them interconnected. Thus we weave in our observations about the anchoring of this three-part vision in the university's experience to date. Finally, we are uniquely tuned in to the opportunity to integrate what we do along the axis from undergraduate to graduate education and research. Our emphasis on connection and integration comes out of our experience at Berkeley as researchers and teachers, seeing the intellectual anchoring of data science in all the parts of the university's mission.

Three dimensions

The stakes in data science vary across the participants. To set out a vision for data science that suits Berkeley, we fill out the three dimensions—deep, broad, and rich—that we outline below. Our guiding thread through these areas is intellectual, as suits a research university, but in nearly every case the implications for practical use in the world of application and industry are profound.[7]

After laying out the categories and identifying key drivers across them, we give substantial detail on each, as seems useful for campus audiences who may not be familiar with all. We pull out the distinctive assets and investments that give Berkeley a unique capacity in each dimension. In particular, we look at the possible hinges that connect them, where we believe some distinctive opportunities lie.

---

[7] In the backdrop are many powerful national-level reports highlighting the profound impact of data science across sectors and domains. We take it as given that data science is a substantive, recognized reality with multiple dimensions. The FAB's task is to map it onto the Berkeley landscape and display a sense of its prospects and needs. One ready reference for its overall significance is the Federal Big Data Research and Development Strategic Plan released by the White House on May 23, 2016.

- **Deep**: A key group of our faculty identify with data science as a core research specialty, one in which they are powerfully intellectually and professionally engaged. In their experience, data science is a proto-discipline with rigorous intellectual foundations, a coherent path of training, and an integrity of its own. At Berkeley a large cluster of these faculty sit in Electrical Engineering and Computer Sciences, in Statistics, or jointly between them. Though this constellation does not encompass every faculty member at Berkeley who identifies with the dimension of depth, it forms a pillar of strength that is unmatched anywhere in the world.

- **Broad**: From the perspective of users across the university, data science is a powerful toolkit of analytics pipelines and data platforms that they apply to advance their own work in a profession or field. In some domains, data science is folded in with other forms of processing, modeling, analysis, and optimization rather than called out as something new and distinct; often this is true of fields that moved long ago to incorporate computation or statistics into their own core. These faculty identify first of all with the "home" field in which data science is put to work, and their concern is advancing knowledge in that domain. Depending on the state of their field, they may be pioneering transformative new methods or straightforwardly using the tools. In some domains they have already formed a coherent identity as practitioners; at Berkeley, computational biology, biostatistics, data-informed public policy, and robotics stand out.

- **Rich**: For a final group of faculty, data science is a bundle of forces at work inside and outside the academy that is transforming societies, human experience, and knowledge. Some of these faculty are technically minded, others are not. All recognize that dealing with the flood of data intelligently and responsibly demands careful, critical engagement, alert to the stakes for the human beings who are its producers, consumers, and subjects all at the same time. At Berkeley we find questions of the human engagements and societal entanglements of data pursued by faculty and graduate students in the School of Information, across a broader landscape of cross-departmental programs and networks, and scattered individuals in many departments and schools.

A common dynamic is operating across these areas. At the decadal scale, data science is an outcome of a reorientation of academic specialties toward information and computation, broadly defined. We can see the establishment of bridge domains like "computational X" and "X-informatics" and the cross-hatching of old fields to create new ones such as statistical machine learning, artificial intelligence, and cognitive science, as well as the emergence of our School of Information out of Berkeley's older library school. The power of a "computational lens" on many disciplines is visible at Berkeley in the wide reach of the Simons Institute for the Theory of Computing. Most visibly for data science, inside some parts of statistics (and mathematics) there are movements to reorient themselves around the intellectual affordances of computing, where new areas have taken powerful hold. Rather than an imperial project coming out of a colonizing center, at Berkeley the shift toward information and computation has been a locally driven process of adoption and matchmaking anchored in the faculty's sense of the future of their domains. It is matched by a dynamic that can now be sensed inside of computer science, at Berkeley and nationally, to reach out even more and form connections with other fields in a way that takes a relationship of partnership and parity as given.

Excursion: The dimensions in detail

Deep
Data science has been built on our exploding capacity to construct high-capacity, flexible, distributed data infrastructures that touch the entire planet and connect to almost every aspect of human, societal, and scientific activity, and to work that data computationally and analytically at (near-)real-time rates and at tremendous scale. At the infrastructural level, a powerful driver in the rise of data science has been our ability to acquire, transform, and integrate massive amounts of data across widely spread systems—bringing with it demands to move that data in part or in whole across distance, to combine its many forms while attending to its real-world messiness and limits, and to query, mine, and protect it, even when it is hard to manage by conventional means. Sophisticated computational, statistical, and mathematical methods are being created atop these infrastructural foundations, extracting significance in alliance with, or independent from, established modeling methods. The domain-by-domain availability of ever-greater masses of data has created urgent pressures on existing analytical methods and technologies. Along with driving the attractiveness of machine learning approaches, it is raising the need for statistically well-founded procedures that provide control over errors, while recognizing uncertainty and enabling transparency into nearly unfathomable complexity. Many bleeding-edge challenges in data science live in this “deep” dimension, often bringing about profound transformations beneath the surface of computational and mathematical platforms on which application domains operate.[8] ● Scalable, robust algorithms and computational, storage, and communications infrastructures to work with immense, noisy, incomplete data and rapidly evolving usage characteristics are frontier areas of research. They present both compelling engineering challenges and foundational theoretical issues. ● The ability to leverage massive datasets will continue to be critical for search, artificial intelligence, speech, vision, and perception, and increasingly underlie interactive systems used by billions of people worldwide. ● Integration of vast, diverse data streams into highly automated control and decision-making systems presents huge opportunities and challenges; this is an area where we must build and study real systems in order to reason about them fully. ● Taking data as a first-class concept has brought powerful new perspectives to the database and software engineering fields. The reinvigoration, even reinvention of this area has created a renaissance in data-centric approaches.

[8] An authoritative and relatively accessible overview from 2013 can be found in the NRC report “Frontiers in Massive Data Analysis,” chaired by Berkeley’s Michael I. Jordan.

- While the web integrated all forms of human-generated information and the many tacit sources derived from digital communication, automated manufacturing, logistics, supply-chain management, and more, it is the case that orders of magnitude wider and deeper streams of "physical" information are emerging through networks of sensors, the Internet of Things, and the deployment of cyber-physical systems, creating fundamentally new challenges and opportunities for management, classification, search, and processing of such non-human generated data.
- The availability of new data flows has created broad research frontiers in the domains that define data analysis. Work on complex analytics in distributed architectures has to keep up with ever-new regimes of speed, responsiveness, and data heterogeneity.
- Analytics has to take on these challenges, critically, under conditions of sparsity or other structured constraints and sharpening demands on the certainty of inference. A key intellectual challenge here is that the assumptions that have underwritten historically familiar analytical methods are apt to be violated in massive datasets.
- Methodologies within statistics are in the midst of being reinvented for the era of the computer, not the hand calculator, when messy data is readily available in mass quantities rather than carefully curated, scarce, and dear.
- High-dimensional statistical inference, optimization in stochastic settings, nonparametric inference, causal inference, and selective inference present many open questions, as do core areas of graph theory and applied linear algebra that are highly relevant to data science. The bridge from computation to statistics and applied mathematics will carry huge intellectual traffic.

Even in areas with major practical ramifications, some of the principled foundations are still being laid. The field is dynamic with significant foundational challenges, as the mathematics to reason from data in an age of pervasive connectivity and computation is simply less advanced than the mathematics to predict continuous behavior unfolding in space and time under physical constraints.

Berkeley has de facto been investing in the "deep" dimension of data science for at least the last 15 years. This has happened through the good judgment of colleagues who, in the faculty FTE process, looked for emerging directions in their disciplines, articulated them critically to campus, kept their eye on what was solid, and bypassed the hype. One result is that between the Departments of Statistics (ranked #2 nationally) and Electrical Engineering and Computer Sciences (tied for #1), we have an incredible concentration of intellectual power and unprecedented joint faculty representation. The presence of a strong faculty group here has consequences across the university as well.

- **In EECS** our strength shows up in shared projects such as collaborative ventures between colleagues in systems and machine learning in the AMP Lab (Algorithms, Machines, and

---

[9] The Statistics External Review Committee's report and the report of the Senate Liaison should be viewed as essential background material for campus decisions regarding data science. In drawing its conclusion about Statistics' readiness to help Berkeley lead, its External Review Committee summarized Berkeley's unparalleled opportunity from the perspective of their discipline. Citing key characteristics for leadership—strength of core disciplines of data science, strength in theoretical foundations based on probability theory, close collaboration of computer science and statistics, strength in the use of data science through multiple disciplines university-wide, and possession by core disciplines of an outward, integrative view that would enable synergies across campus—it observes, "While other institutions can claim some of these 5 characteristics, we know of none that combines them all."

People), a technical launch-point of the Presidential Big Data Initiative and a font of widely adopted and commercialized open-source technologies, and in a powerful group in artificial intelligence, which has charted a Berkeley path through the field with a strong vision and deep historical roots, as well as in databases, distributed systems, controls, optimization, and signal processing. With the creation of the Simons Institute and its programs on the foundations of data science, a bridge to deep issues of theory has been built. The effect carries forward as new faculty join the campus and find their research prospects transformed by opportunities for collaboration in the cultural context of Berkeley EECS.

- **In Statistics** Berkeley is known to be unique in the world for the strength and confidence of our faculty in shaping their future in an era of massive data. Our statisticians' position is grounded in their depth in statistical theory and probability theory, along with the long-standing fluidity with which they work with colleagues in computer science, signal processing, and information theory, as well as colleagues in domain areas such as genomics and neuroscience. At present, the boundaries are so fluid that more than half of our Statistics department has some form of faculty appointment with other departments, including a large number directly with EECS. The External Review Committee for the department's 2015 Academic Program Review called out the department's and Berkeley's truly distinctive situation, concluding that "the Department of Statistics is ready to help the University establish itself as global leader in data science."[9]

- **In departments outside of EECS and Statistics**, faculty with a "deep" data science identity are being hired as well. As examples, the Department of Industrial Engineering and Operations Research and the School of Information have standing strength in this area and have devoted searches to it in the last several years.

This pillar of data science strength, which is spread collaboratively across outstanding departments, is noticed outside of Berkeley as well. Along with real-world impact (natural language processing that is now pervasive with Siri and others, Apache Spark that underlies most large commercial data analytics, image-based recognition and 3D reconstruction, high dimensional statistics and graphical models used in social networks and genomics, to name a few), Berkeley faculty have trained a host of outstanding PhD students who been sought out to shape data science programs elsewhere.

More dangerously, Berkeley has been losing key faculty piecewise. Over the last several years we have risked becoming the data science poaching ground for other universities seeking top-level leadership for their initiatives. We staved off this dynamic in multiple cases up through 2014-15, but in 2015-16 we failed. Until Berkeley can take concerted action on a data science initiative, we can expect losses of greater magnitude to occur.

Broad

The "broad" dimension of data science is its platform of methods and tools that get used across diverse fields. In academia, massive data and computation have already worked their way into fields ranging from high-energy physics, bioinformatics, and astronomy to environmental sciences, civil engineering, neuroscience, linguistics, and natural history, to take only some local areas of strength. The list goes on: data science is opening new opportunities in the social and behavioral sciences and the digital humanities, leaving it hard to see what areas of research will be left entirely untouched. The same opportunities opened up by new datasets can be found in many of the occupations for which we prepare professional students: business, public policy, urban planning, public health, and education, as well as the career path of data scientist itself.

While the specifics vary field by field, the availability of data is common across them, as are some of the broad-brush opportunities and many of the challenges.[10] As application-domain practitioners pick up data science methods and tools, the availability of new caches of data and advanced analytical methods has begun reframing the practice of research in ways that show commonalities across disciplines and fields:

- Along with generating highly focused observations under controlled conditions, researchers can move to **scanning large datasets for knowledge buried within them**. Inferential techniques such as those developed for astronomy, cosmology, and genomics let us work out in the long tails of the distribution. In needle-in-a-haystack situations, we can pick up on the rare phenomena, rather than just the bulk, and design interventions that target individuals for differential treatments or recommendations, as promised by precision medicine, for instance. Data analytics can let us pick out subtle patterns in corpora, whether in digitized archives of specimens and associated data in natural history collections, textual analysis of literary materials, network or temporal analysis of social media data, or machine-learning models of cognition, vision, and hearing. The availability of large training sets is part of what has made it possible for computers to understand what we say or recognize what we see. At the same time, challenges around interpretability—what does it mean that a pattern can be observed? that behavior can be predicted?—can raise deep questions about integration with theory, practice, and policy. In some areas, still, such as domains where creation and experimentation with new materials is the order of the day (materials science, chemistry, some forms of biomaterial design), the end-run around theory afforded by large databases may be the way forward. In effect, we can embrace dimensionality rather than eliminate it, working more fluidly in areas that are less under our prescriptive control.

- Data science creates opportunities to **systematically integrate datasets with each other and with computational models**. Instances of gaining new insights by marrying disparate sources of data range from neuroscience studies combining imaging data with medical records and with genomic information, for example, to social science investigations that exploit partial linkages between administrative and commercial databases and align both to individual-level records of behavior. Explicitly complementary kinds of evidence can be brought together to get at systemic phenomena, as when longitudinal studies in public health (repeated observations of the same variables for individuals over time) can be brought into new lines of dialogue with cross-sectional studies (of a broad study population), bringing in many more confounders that are in fact conditioners to be drawn out of the data. In deeply transdisciplinary collaborations, it

[10] It is hardly technological determinism to see a key driver here in the availability of data. Wide-ranging changes are being advanced by data streams from high-resolution or high-throughput facilities for observation and laboratory measurement, automated instrumentation of processes and record-keeping, high-density or widely spread networks of sensors and devices, massive quantities of information now available in audio and image formats, rapid growth in collections, corpora, and streams of deliberately digitized or born-digital data, even large-scale, high-resolution simulations.

is now possible to integrate historic data from biological specimens, accompanied as they are by field notes and ecological data, with climate sensors, remote sensing, landscape layers, and genomic data, in order to infer how organisms will adapt or respond to climate or other modifications of the environment. Multi-scale, multi-fidelity modeling can be done for atmospheric or terrestrial phenomena all the way down to tiny sensors, all the way up to remote sensing, and all the way across to remote parts of the world. It has become possible to ingest enormous amounts of data into sophisticated models that already exist, as in weather prediction and climate.

- The capacity to **use ongoing data acquisition to shape the model itself, and the world**, is powerful across disciplines, from urban transportation planning to control theory. Rather than truth once and for all, data-driven models may be focusing our attention on an evolving truth that is shaped by the processes of data collection and intervention itself. That integration can be deeply characteristic of data-intensive work that is indeed aimed not only at observing, but directly at changing the world around us, as motivates much cutting-edge data science work in professional contexts and schools.

- Ultimately, **new modalities of data collection give us access to important phenomena** we could only inch our way toward before now. Being able to sequence the collective genome of species' microbiota (the microbial flora found inside organisms) creates a wealth of information that was previously unavailable in specificity or volume. Microbiome research touches multiple areas in the biomedical, environmental, agricultural, and evolutionary biology communities. In the social realm, digital exhaust from online platforms can give information about human decision-making or network dynamics; collecting information via people's mobile devices or other instruments of personal tracking gives us access to behavior "in the wild," not just in the lab. With observation and experimentation on social media, there are effects looping back to shape the world we are registering, underwritten by our new channels of connection and communication that are also our research instruments. Exactly here, with human beings in the loop, the rich complexity of the conduct of data science research comes immediately into view.

Success in the "broad" dimension of data science has depended on drawing in experts with multiple

[11] We did not have access to departmental proposals for faculty FTE. For some understanding of departmental strategic planning, we have drawn on information from colleagues and responses to a survey of department chairs and deans.

[12] Computational biology was called out in 2002 as one of the campus's strategic initiatives, leading to a graduate group, a PhD program, a Designated Emphasis, and a center. The Center for Computational Biology is one of the four New Initiative Centers (NICs) that had been reporting to the Vice Provost for Strategic Academic and Facilities Planning. NICs do not hold faculty FTE (faculty have homes in departments instead), but faculty who are hired through a search process involving a NIC dedicate 50% of their teaching and service to its mission. In 2015-16 the NIC model was acknowledged to be in need of rethinking and the NICs were asked to explore possibilities for new decanal homes. CCB is one of two NICs (the other being the Berkeley Center for New Media) for which a decision was deferred until 2016-17, after the completion of the Data Science Planning Initiative.

[13] Designated Emphases: Computational Biology; Communication, Computation, and Statistics; Computational and Data Science and Engineering. Masters offerings: Masters in Information and Data Science (online), Biostatistics, Statistics, track within the Haas School of Business MBA, two Masters of Engineering programs (IEOR, EECS). Graduate traineeships: DS421, Environment and Society: Data Science for the 21st Century, funded by NSF; a new Biomedical Big Data Training Grant, funded by NIH.

competencies to build models, methods, algorithms, or systems tuned to particular classes of problems. It is hardly the case that data science just replaces existing approaches. It feeds into them, extends them, complements them, sometimes conflicts with them, and the plurality of relevant methods and the context-specificity of domain-area questions shapes the ways its techniques can get used. Outside the university, too, in business, civic, and policy settings, research design is understood to be a critical piece in applying data science. That is, context matters immensely for understanding the data and the conclusions drawn from it, backing them up against the existing base of knowledge, theory, and questions. This appreciation is a hallmark not only of serious research, but also of the educational programs that Berkeley has developed in data science, from our professional Masters in Information and Data Science to our entry-level undergraduate Foundations of Data Science class and the "connector" courses that relate to it symmetrically.

As with the "deep" dimension, much of Berkeley's development in "broad" data science has happened from the ground up in a distributed fashion. It is not uniform or pervasive, even if it is a trend. Along with concentrations in recognized areas such as biostatistics and data-informed public policy, the information we have points to it as a growing element in the portfolio of many departments. It is becoming an ever more important part of professional practice and professional education across multiple schools and colleges, making itself felt in a diversity of ways. It shows up in the incremental retooling of faculty research programs, the hiring of junior faculty, and new undergraduate and graduate courses being offered by faculty brought in to represent data science methods, approaches, and tools.[11] We note some distinctive features:

- Recent years have seen the **emergence of new cross-departmental areas** of computational or data-enabled research and teaching, as particular clusters of disciplines evolve on their own to bring these opportunities to the fore. Strikingly, these areas appear in diverse parts of campus. The rapid crystallization of Digital Humanities at Berkeley and the emergence of a cross-departmental cluster in robotics are conspicuous examples of the last few years. The "datafication" of much work at the crossroads of environmental design and engineering is ongoing. On the methodological side, the broad reach of Berkeley's applied statisticians has been instrumental in creating productive new collaborations (for instance, with environmental science and neuroscience) and seeding a spirit of openness to cooperative exploration and partnership.

- Berkeley's one targeted investment in "broad" data science has been in **computational biology**, started nearly fifteen years ago.[12] The impact of our computational biology program shows the effect that a strategic experiment can have—and the profound challenge faced by any program that lacks control over resources and a strong institutional home. It is safe to say that Berkeley's earliest strategic investment in a data science-related area is at a crisis. Disillusionment with campus's ability to support the Center for Computational Biology has led top researchers to draw back from its governance, go on the job market, or leave Berkeley entirely.

- Many institution-wide experiences point to the emerging importance of **"broad" data science across the Berkeley landscape**. These include the connector outreach and uptake of our undergraduate data science curriculum, the emergence of Designated Emphases (graduate minors) and data science-inflected masters offerings, and our success in winning exciting externally funded graduate traineeship programs.[13] Responding to the demand for "broad" data science, Berkeley has created important regional research-supporting platforms such as D-Lab (the Social Sciences Data Laboratory), the Computational Genomics Resource Laboratory, and the Geospatial Innovation Facility, invested locally in data-intensive facilities in departments and across colleges and schools, and supported new campus-wide offerings from Berkeley Research Computing that reach into this space. In recent years, the instinct for sharing experience across disciplines has made itself felt in the shape of the **Berkeley Institute for Data Science**, about which we say more below.

Ultimately, “broad” data science may become pervasive, indeed, ubiquitous across many departments. At that point we would have a very different university landscape.

Rich

Data science has stakes, finally, for the human beings who are its producers, consumers, and subjects. The datafication of human experience is being pushed forward at huge speeds, driven by motivations ranging from exploring intellectual and technological frontiers to making lives better through data-driven services to creating massive concentrations of wealth, power, and control. Rapid-fire questions around the societal, normative, and user-facing aspects of data science emerge at the same time, often pushed forward by examples that display where we are in uncharted terrain.

- New forms of data pry open **ethical and societal questions around humans as subjects** in a data-driven world. These questions shape research design, sampling strategies, and privacy-preserving analytical methods; they touch on practices of sharing of datasets and the effects of the research we do back on the world we observe and engage. Even as government and corporations collect massive amounts of data about us, debates about ownership, consent, privacy, and surveillance are being tackled with inherited conceptual and policy frameworks. The same is true for the implications of machine learning and AI for structures of societal accountability. Algorithmic decision-making can work against discrimination or reinscribe it, and the differential exposure of social groups to advanced analytics can create disproportionate advantages (access to advanced medical treatments) or exposure to mechanisms of societal control (DNA registries in relation to the carceral state). The emerging field of data and ethics takes up normative questions at all levels of analysis and does not shy away from talking about interests and power.

- In a data-rich world, **human decision-making and sense-making** takes on new forms. Medical choices about how to engage with the outcomes of data analytics are already upon us; behavioral self-monitoring with mobile devices and activity trackers is as well. As personal decisions accumulate, they move from the terrain of the individual to that of the social, as knowledge and choices about the self bring with them implications for others (family members, peer groups, users of shared platforms and services). Moving the needle with data inside organizations or societal contexts takes far more than just presenting seemingly neutral facts. Human beings are believers, not just sources of data. In a Bayesian epistemology we face deep questions about how human beings perceive different elements of the world and still converge (if they do). Human interpretation and sense-making intrinsically come into play, too, in ways that pull in media and the arts. Some of the areas in which the significance of data science will be most profoundly explored will be disciplines where it evokes methodological conflict, particularly the humanities and the interpretative social sciences. And on the side of the natural sciences, there are methodological challenges in operating in a data-driven world that may be effective at robust prediction but short on fundamental, principled understanding.

- **Facilitating human users' interactions with data** is a major subject of inquiry in the emerging field of human-centered data science. This area includes studies of human-computer interaction (HCI), data visualization, exploratory data analysis, and end-to-end pipelines with humans built in. Data science in the "broad" dimension is part of a value-laden reconfiguration of scientific practice around the sharing of data and code. It calls on collaborative platforms, reproducible practices, and computational data narratives (as in notebooks) that are increasingly the subject of deliberate design.

Grasping the human aspects of data science is a demand we have heard from industry contacts no less than academic colleagues. In the last few years, a particularly urgent sense around data and ethics has been striking to see. This need can be met by means of curricular exposure and cross-training, or by intellectually challenging options that embed practitioners of "rich" data science in the midst of operational teams. The "rich" dimension can be highly technical, as in differential privacy, fairness-aware data mining, certain lines of research in computer security, and a diverse body of work drawing on behavioral science. The older model of separate efforts to tackle the ethical, legal, and social implications (ELSI) of science and technology, conceived as distinct from the technical, has been yielding ground in industry and in academia to more complex, more difficult, but often more effective forms of engaged collaboration and critical technical practice.

Given our university's public mission, an essential element of data science at Berkeley lies along the "rich" dimension. In particular, attention to the ethical dilemmas and human entanglements of data analytics and pervasive computation is central to our identity and to our public service. This spirit animates multiple parts of our activity, including education no less than research; societal and ethical considerations appear our growing data science curriculum starting at the foundational level. A distinctive aspect of Berkeley is that we do nothold back from talking about the social and political constellations in which data science is unfolding. This, too, is true on the nominally technical side (if that distinction makes sense any more).

- Berkeley's capacity in the "rich" dimension has a historical anchor in the **School of Information.** The social, organizational, and technical aspects of information are core subjects of attention, and centers and clinics bring together faculty, researchers, and professional and PhD students around thematic areas including information law and policy, cybersecurity, and technology, society, and policy. One of the commitments of the I School is its insistence on cross-connecting the technical and the social, bringing both lenses to bear on problems of information systems and design.

- In **engineering domains**, the College of Engineering has world-leading areas of strength in data security and privacy. Courses, programs, and student groups pursue questions of engineering and ethics, and data science-related efforts along the "rich" dimension have been integrated into CITRIS (the Center for Information Technology Research in the Interest of Society) and the Fung Institute.

- Cross-departmental faculty and student networks in **science and technology studies** in the humanities and social sciences have picked up this area with projects on algorithms and data in the Center for Science, Technology, Medicine, and Society. BIDS has invested in an in-house research team in data science studies that shapes international discourse in the field.

- Attention to **design and the artistic and humanistic aspects of data** has emerged in different corners of the campus, including the Berkeley Center for New Media, Digital Humanities at Berkeley, the Berkeley Institute of Design, and the Jacobs Institute for Design Innovation. As data becomes more a part of our everyday landscape, we can expect this productive dispersion and engagement to expand.

Though cross-campus discussions on the “rich” dimension of data science will generate differences of opinion, they will be necessary, appropriate, and important to have. Figuring out what will supplant our established frameworks for engaging with data is about as contested a discursive field, and about as embedded in large-scale debates about societal transformations and political economy, as intellectual property became as the Internet scaled.

Hinges and connections

Depth, breadth, and richness are not cleanly separable. Much data science work touches on more dimension than one. In fact, while a great deal of “broad” data science applies relatively standardized methods and techniques, at the cutting edge of the field it is often in need of exchange with “deep” areas. Two-way traffic can lead to more robust appreciation of constraints and trade-offs, better understanding of methodological limits and failure modes, and collaborative exploration of extensions and new approaches. Faculty with joint appointments across the boundaries (between Political Science and Statistics, for instance, or Computer Science and Molecular and Cell Biology) exemplify this aspect, as do patterns of project- and grant-based collaboration (between neuroscientists and statisticians between or electrical engineers and cognitive scientists, to take Berkeley examples that have brought powerful insights over the long term).[14]

Historically, statisticians and applied mathematicians have pioneered this kind of interchange with application domain areas, collaborating on real problems with fresh data to generate new methodological insights. They may stick with one application domain or move between several, as their mathematical and methodological interests suggest. Computer scientists can make similar moves, seeking out motivating models or test cases for principles they are seeking to explore. In computational biology and a few other well-established “computational X” areas, they have become full partners with domain-area researchers. For this partnership to work, it bears saying, the problems have to be at the research frontier in both disciplines; pure service roles do not attract faculty of Berkeley’s caliber. Nor can the relationship be engineered for the sake of institutional convenience. The draw for Berkeley faculty will always be the payoff in developing of new methods, technologies, and knowledge.

---

[14] We have scraped campus databases and reviewed joint appointments, faculty research interests, and grant data as far as it has been made available to us. A team of students from our data science education program continues exploring these datasets. Making Berkeley legible to itself is a key need that many other programs on campus are investing in as well.

Similar sets of hinges connect “deep” to “rich” and “rich” to “broad”—for instance, around causal inference in relation to policy formation in domains where the intelligibility of decision-making must be clear, privacy-preserving systems and algorithms for strong use cases under particular regulatory rules, or multi-phase challenges around integrating human judgment into analytics pipelines. To do this work well requires bringing credentials and expertise from more than one side. At the other end, for many challenges along the “rich” dimension of data science, it is critical that its practitioners *not* stand entirely apart from the others if they are to understand their problem domain from the inside.

Data science is marked by the near-fractal nature of the disciplinary boundaries within it. The interdigitation of practitioners coming from different domains is part and parcel of the field. In this way it is characteristic of a larger set of processes in research across disciplines. Its hinges and connections are a deep reason why data science so often troubles conventional practices by which universities place faculty within programs, departments, and schools.

Observations

It would be possible to take on just one or two of the aspects of this deep, broad, and rich vision. Many of our peers have followed this path. We think the unique opportunity at Berkeley is to strengthen the whole by building in the interconnections, both anchoring them within a definite organizational form and reaching across porous borders to stretch out campus-wide. This sense comes from three places. First, the multidisciplinary reach of Berkeley’s excellence means we have strength to build on across the board. Second, the intellectual breadth and collaborative openness of many of Berkeley’s “deep” data science faculty makes them receptive to a more “broad” and “rich” venture than their colleagues elsewhere. This is happening at a time when our faculty in “broad” and “rich” areas are simultaneously proving less invested in maintaining older forms of differentiation. Finally, our experience in the last five years with campus-crossing organizational experimentation has shown us how effective we can be when working together. That last element of our trajectory gets focused attention in the section that follows.

This vision does not mean that Berkeley can, or should, do all of everything. It does give us a star to steer by. Taking it on is a challenge that will require strategic planning, careful development of selection criteria for our options, and unusually thoughtful leadership that must cultivate and reward collaboration and openness in a new way. If we move forward on this path, we believe Berkeley can pioneer a globally unique integration across the dimensions of data science, as far as it is feasible with an organizational design that can deliver on the promise and compatible with the resources and mechanisms we can bring to bear.

3. Campus-crossing Organizational Experimentation for Data Science: Cases and Lessons

The last five years have seen campus-crossing organizational experimentation around data science, at Berkeley no less than at other universities. How to facilitate these experiments is an important challenge for our campus to address, and we must take it on with thoughtfulness and realism. At Berkeley, early data science-related instances of campus-level, interdisciplinary efforts include the Simons Institute for the Theory of Computing[15] and D-Lab, the Social Sciences Data Laboratory.[16] Both have relied heavily on connectivity across campus, on a culture of openness and outreach, and to some degree on collective governance.[17] More specifically, the opportunities distinctive to data science have led to two significant campus-level experiments, the Berkeley Institute for Data Science and our undergraduate data science curriculum. Because the FAB's thinking about organization, mechanisms, and culture has been influenced by these two prototypes, we describe them briefly and identify lessons we think can be drawn.

---

[15] The Simons Institute for the Theory of Computing opened in July 2012 with a five-year grant of $60M from the Simons Foundation. It reports to the Vice Chancellor for Research and is housed in Calvin Hall. Its mission is to bring together the world's leading researchers and scholars in theoretical computer science and related fields, expanding the horizons of the field by exploring other scientific disciplines through a computational lens. Its major mechanism is a series of semester-long programs with long-term participants, joined by short-term visitors attending workshops during the semester. The Simons Institute has been experienced as a boon for faculty at Berkeley in an impressively wide range of scientific areas through its programs and the semester-long local presence of the researchers it brings in.

[16] D-Lab opened in February 2013 as the outcome of a visioning process collaboratively spearheaded by the Vice Chancellor for Research and the Dean of Social Sciences in the College of Letters & Science. Its aim is to empower Berkeley researchers with methods, tools, and services for data-intensive social science. Because some of these capacities apply outside the social sciences as well, D-Lab's training and consulting offerings are de facto used by researchers from other parts of campus. D-Lab initially reported to the Vice Chancellor for Research and now reports to the Dean of Social Sciences. Located in Barrows Hall, it is supported with important investments of decanal and campus resources and has drawn some of its strength from cooperation with deans of other colleges and schools.

[17] Other, more localized arrangements have also informed our thinking. These quite often have been catalytic in energizing collaboration and advancing research. They include experiments at widely varying scales that have been housed quite comfortably in one general disciplinary area, from the Berkeley Initiative for Transparency in the Social Sciences to the California Institute for Quantitative Biosciences (QB3-Berkeley), and others situated in a space that brings together a very broad range of disciplines, such as the Berkeley Initiative for Global Change Biology. One common thread is a strong sense of mission and collaboration. Another is the ability to turn small investments of campus resources into significant outside support.

The Berkeley Institute for Data Science

The aim of BIDS is to advance data science in application to scientific and social scientific disciplines and make the university more hospitable to its practice. It provides a common ground across disciplines where shared tools can serve as a bridge to foster interdisciplinary research and where ideas and analytical frameworks can be discussed with those familiar with the appropriate tools.[18] With that breadth of mission, BIDS's early definitional process began in Fall 2012 and taught the campus about the range of potential engagement, drawing in several hundred researchers who came together for its brainstorming sessions, weighed in on its proposal process, or assembled to mark its Fall 2013 launch. BIDS was created with a five-year investment of philanthropic funding from the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation, with a commitment of campus resources appearing in the form of faculty time and a generous agreement by the Library to make available space in Doe Library for this duration.

BIDS does significant service to the university as a shared space in the center of campus that is available for diverse participants to gather and discuss research, with a broad reach across campus research disciplines and an interest in integrating the full academic spectrum, including faculty, postdocs, graduate students, and undergraduates. One of its goals is to provide a common platform for researchers to come together across domain areas on campus and at Lawrence Berkeley National Laboratory—to take a recent example, exploring shared challenges in image processing across domains ("ImageXD"). It is developing a new incubator ("Machine Shop") to partner with postdocs in labs and undergraduate students in the URAP program to build software tools for science.

In Machine Shop, integration with faculty-led research is being piloted in an exploratory way, through common projects in the service of research needs with contributions from both sides. This is the outcome of a process of learning from BIDS's first two years of experience, as it became clear what forms of connectivity to reach faculty might be easier than others. Other models are inviting as ways of generating connectivity and network effects, including faculty co-mentorship of fellows and other BIDS participants.

BIDS has strength in collaborative open-source tools and platforms in the service of data science, particularly in open and reproducible practice. In the national data science landscape it is widely known as one of the bases for Project Jupyter, a multimillion-dollar effort that develops open-source software for interactive and exploratory computing; rOpenSci, a broad-based software collective that provides R-based tools in support of reproducible workflows; and computational fellows who are core developers of the major Python libraries NumPy and scikit-image. Locally it is a convening space for groups of

---

[18] BIDS is a research unit reporting to the Vice Chancellor for Research. It was launched as a new organization in December 2013 and opened its doors in Doe Library in August 2014. It is headed by a faculty director and has a senior fellow cohort of faculty affiliates and LBNL researchers. It employs research and administrative staff and houses postdoctoral and graduate data science fellows. While it supports informal training in data science, it does not have responsibility for undergraduate or graduate education or hold faculty FTE. Its initial funding from the Moore and Sloan Foundations amounts to roughly $13M for five years. It partners with two other institutions, the eScience Institute at the University of Washington and the Center for Data Science at New York University, within those foundations' Data Science Environments program.

students interested in data science who have used it to host events and participate in project-based learning outside the classroom, as in the BIDS Collaborative.[19]

In addition to developing new cutting-edge tools, there is an opening for BIDS to explore how to support the use of existing tools to answer similar questions across diverse data and explore the commonalities across disciplines, as in the case of ImageXD. This form of support to science application domains requires particular forms of outreach directly in the service of what may be seemingly routine research needs. Other units on campus may be able to contribute to this effort, as suggested by the emergence of consultant networks, Berkeley Research Computing, and units like the Computational Genomics Research Laboratory, the Geospatial Innovation Facility, and D-Lab.

BIDS also serves a platform for varied groups that reach across campus in the data science space. In the area of practical training for data science, it carries responsibility for the local institution of the Python Bootcamp and provides other training offerings. It will host a cross-campus summit of programs, departments, and support units involved in graduate-level education in data science in early Fall 2016, together with the Data Science Planning Initiative.

BIDS has been working hand-in-hand with the Data Science Education Program in the area of education and training. Several of its data science fellows are involved in supporting or teaching in the data science curriculum, and BIDS has provided space for instructor office hours and student meetings. The teaching platform for the curriculum is built around Jupyter notebook technology, and the contributions of the Jupyter team in BIDS and beyond have been absolutely essential to the capacity of the curriculum to provide its computational environment at scale.

The Berkeley data science curriculum

Berkeley's data science curriculum is a faculty-led effort engaging a broad range of campus departments in offering classes that meet student demand for data science. In its first year of operations, 2015-16, it began implementing a staged plan sketched in 2014-15 by the Data Science Education Rapid Action Team (DSERAT), an eight-faculty-member committee that consulted broadly across campus in order to create a draft curriculum design. The curriculum responds to the groundswell of interest among Berkeley undergraduates in data-science related coursework, matching the sense of significant numbers of faculty that critical engagement with data is of key importance to many fields of study, central to job opportunities in many industries, and integral to personal and professional decision-making in any walk of life.

The design for Berkeley's curriculum is a multi-tiered program that is structured to address the breadth, depth, and richness of data science. Its entry-level offerings, which also support Berkeley's Undergraduate Initiative, have attracted significant attention at other universities, both for their conceptual content merging computation and inference and for the core-and-connections program architecture. The program is welcoming to students of broadly spread interests and backgrounds, reaching across Berkeley's diverse student body from technical majors to the social sciences and humanities. It is built on the assumption that a data science

[19] The BIDS annual report can be consulted for more information.

education program for Berkeley ought to speak to the social implications of the spread of data into all spheres of life.[20]

- The curriculum is anchored in an **entry-level class, Foundations of Data Science**, cross-listed between Computer Science, the School of Information, and Statistics, and collaboratively taught (CS/Info/Stat C8). The class presents key elements of computational and inferential thinking in an integrated fashion. It uses the affordances of computing to engage students in manipulating real data, offering conceptual understanding through hands-on work that introduces societal, ethical, and interpretative issues in a fluent, contextual way.
- Closely allied with the Foundations course is a suite of **entry-level "connector" classes in other departments**, engaging many disciplinary areas in line with students' interests and prospective fields of study.
- The entry-level offerings now satisfy **prerequisites or requirements in many departments** and count for the Letters & Science Quantitative Reasoning requirement. **The program is not a university requirement.**
- The goal is to **build a comprehensive curriculum starting from the bottom and connecting across the university**, beginning with this set of introductory offerings that can make data science accessible to any student at Berkeley at the same time as providing a high level of conceptual understanding. The entry-level courses can provide the base for later classes in a broad range of departments that will be able to leverage and extend what students have learned.

The classes in the Data Science Education Program are new. They have been approved by the Academic Senate's Committee on Courses of Instruction. In 2015-16 the Foundations of Data Science was taken by roughly 550 students (fall pilot and spring regular offering), and 12 connector classes were offered on topics such as smart cities, cognitive science, literature and data, race and policing, ecology and environment, geospatial data, data and ethics, statistical foundations, and computer science foundations. In 2016-17 the entry-level part of the program will continue at the level of roughly 500 students in the Foundations class per semester and a similar number of connector offerings. Courses now in the design stage include advanced classes in data science that, if approved, can become part of an eventual data science major, as well as minors that can integrate back with students' (other) major programs of study. Further, they include follow-on courses and data science modules that are being developed in a growing group of other major programs.

Through the experiment of creating a new program in this ground-up way, the campus has been able to learn about the needs of students, faculty, and programs. For instance, the DSEP was able to survey students in the Foundations of Data Science class in Spring 2016. Over 50 majors or intended majors

[20] The Data Science Education Program (growing out of the DSERAT report) has been steered to date by faculty who serve in the leadership of the Data Science Planning Initiative and has operated from an interim home provided by the L&S Dean of Undergraduate Studies. It was launched with faculty and departmental contributions in time and in kind, substantial assistance from the L&S Undergraduate Dean's staff and reserves, temporary support from a campus allocation of TAS and administrative funds, and fundraising and donations assisted by UDAR, the Department of Electrical Engineering and Computer Sciences, and the College of Letters & Science.

were represented; 42% of respondents identified as female, 58% as male. When asked at the end of the semester about how interested they were in pursuing a potential data science course of study, on a scale of 1 to 5 more than 50% of them indicated an interest level of 4 or 5 in a data science major; more than 75% indicated that level of interest in a minor (418 responses).[21] Now that the curriculum is becoming real, the DSEP has also been able to work with a team of undergraduate student researchers in BIDS to investigate the reactions of diverse populations of students across campus to data science, and to use those findings to shape the program's offerings and outreach strategies.[22] Student interest is documented in other ways as well. Faculty will now need to look for effects on students' academic careers and performance moving forward.

From interactions with current or self-identified potential connector instructors,[23] early indicators suggest that the curriculum's core-and-connections architecture and facilitation help create desired effects of intellectual integration, inclusiveness, bi-directional influence, and network effects. These effects are necessary for the DSEP to be able to tune its approach, particularly to identify to what extent it can support upper-division learning for Berkeley students who pass through the entry level and pursue a range of other majors across campus. Ensuring connectivity and community is not trivial, and there are significant parts of campus whose needs are not yet being met by the program. As initial feedback, this indicates that important work is still to be done to see how far such bridging is possible.

Reflections

Berkeley has been the venue of many co-located experiments in data science. These are not exhausted by the two examples described above. In reality, the Berkeley vision for data science is already being tried out in a broad range of programs in different settings, programs that are incrementally finding ways to work together. This ground-up approach has allowed the FAB to lay out a Berkeley vision for data science in a collaborative, transdisciplinary way. At the same time, we see that much of the campus-crossing organizational experimentation in data science has been carried out in an interstitial space between academic units. This is a powerful way to nucleate new activity, but it has organizational costs and a limited lifetime for sustainability.

---

[21] The DSERAT report advanced the hypothesis that a substantial component of the massive growth in recent years in computer science course-taking (in student enrollments and majors) might in fact indicate interest in areas that could be better covered in data science. As computer science course-taking continues to grow—this year it surpassed mathematics in student credit hours for the first time—this is a hypothesis that the data science curriculum will allow us to test. More importantly, if the hypothesis is correct and we take action on that finding, it will allow us to better meet student needs, which is what matters at the end of the day.

[22] The four student teams supporting the DSEP are Infrastructure, Outreach, Diversity and Inclusion, and Mapping.

[23] In June 2016 the DSEP ran an experimental week-long course on data science pedagogy and practice for potential instructors, introducing them to materials and approaches from the Foundations course and connectors and supporting them in exploring data science approaches for use in their own teaching. 70 potential instructors applied, of whom roughly half were able to attend, coming from programs ranging from Math, Physics, and Economics to Neuroscience and Linguistics to Near Eastern Studies and American Studies. Feedback was strongly positive. Outcomes include new connectors and course modules that are now being implemented and plans for convening a larger community of practice.

Data science is just at the start of its evolution, and much of it will remain deeply embedded in home units across the campus who will want to pursue it in their own ways. Formal and informal mechanisms of horizontal collaboration between disciplinarily organized academic units are continuously under development, so we can both draw from past experience and acknowledge that there is much we still need to learn. In the context of campus-crossing organizational forms, it is natural that our thinking has evolved through these experiments.[24] In particular, BIDS's ideal of an open interdisciplinary community on neutral territory and the architecture and culture of the data science curriculum, suitably generalized, have come to serve as powerful prototypes for the FAB's thinking.

[24] The membership of the FAB includes several faculty from the BIDS co-investigator team and all the participants on the 2014 DSERAT, now strengthened by a roughly equal number of faculty in broader domains. Michael Franklin (EECS) participated in the DSERAT and was initially appointed to the FAB. In 2016 he was recruited to the University of Chicago to lead their data science initiative.

4. Expectations of an Initiative in Data Science

If our aim is to set up Berkeley to lead this field for the next decades, we need structures and approaches that learn from and go beyond our piecemeal strategy so far. Framed as four expectations to be met by any solution, we want to:

1. Maintain and strengthen **faculty excellence**.
    - Create a pillar of strength in the core of data science and its ramifications, building in practices and structures that support and grow these fields.
    - Do so while accommodating the anchoring of data science in world-leading existing programs with strong traditions that extend beyond data science.
    - Amplify the benefits of intellectual synergies in these fields by encouraging collaboration across them, supporting them culturally and practically.
    - Connect them to domains of application that strengthen them bilaterally, and support application domains as they move in data science-enabled directions.
    - Have the capacity to build areas of faculty strength in strategically identified application domains, and exercise it if we choose.

2. Offer **outstanding education** at scale.
    - Provide strong programs from undergraduate to professional to doctoral levels, with pathways to jobs and top research careers in an increasingly data-centered job market and research orientation.
    - Deliver on the responsibilities of a large, multidisciplinary education program that many units across campus can be expected to depend upon.

3. Develop **strategic organizational capacities** to allow Berkeley to prosper in this area.
    - Have mechanisms and practices at our disposal that let us readjust the diverse fields involved in data science and their interrelationships, strategically shaping them together organically over time.
    - Develop and exercise the strategic capacity to keep the campus as a whole in a winning position in data science many moves into the century ahead.
    - Implement strong, successful programs for revenue generation, both in professional education and in research, and attract major philanthropic support.
    - Grow robust linkages to the industry ecosystem, capitalizing on our location next door to the world's leading engines of innovation in data science.

4. Embody a **distinctive Berkeley vision for data science**.
    - Develop strong, integrated programs that bring together the three dimensions of deep, broad, and rich data science.
    - Project this vision into the world as Berkeley's unique contribution.

    - Ensure that vision aligns with Berkeley's core public mission, emphasizing data science's ability to promote the public good in its multiple domains.

We could leave these as goals to be worked toward by existing practices and mechanisms. However, business-as-usual approaches often produce local optimizations instead of global ones. They move slowly and incrementally and, for a set of goals of this ambition, have a fair chance of failing. For completeness we also diagnose some limitations of current approaches:

- We face considerable challenges in meeting the needs of faculty and departments at the core of data science.
    - We encounter substantial friction and coordination costs in trying to do strategic planning (for educational programming, resourcing, fundraising, or faculty FTE) for academic units whose futures appear to be deeply intertwined while they report to multiple deans.
    - We fall short, conversely, at supporting these programs at the campus level when each sits within a larger constellation of units from which their individual profiles and paths diverge.
- We do not do justice to key areas that would flourish in alliance with those fields.
    - We have not yet developed solid organizational forms to support proto-disciplines that have grown up in the penumbra of computing and data, or are doing so now.
    - We lack strong enough connections from foundational data science disciplines to key parts of the academic landscape where research frontiers are emerging. At present the quantitative and computational social sciences stand out.
- We fall short of the strength we could achieve through integration and alignment. In their absence we create risks for our success.
    - There are missed opportunities in the campus's broadly spread portfolio of revenue-generating professional education programs when we cannot call on the capacity to build between core departments in this space.
    - We have a segmented academic environment with a research center and an education program with huge potential in this cutting-edge field which are as yet unconnected organizationally to the academic units that need to sustain them.
- We let incremental decision-making hold us back from progressing.
    - If we continue in our present holding pattern, we will begin losing the advantage we now hold. In truth, several key recent losses of faculty show we already have.
    - We have the potential to do far better on philanthropy and revenue generation, given our assets, reputation, and historical ties.

5. The Faculty Advisory Board's Recommendations

The FAB has concluded that the university needs to undertake a deliberate set of strategic actions, placing responsibilities at both the campus level and among the faculty concerned, to shape this future by coherently planning, fundraising, and allocating resources at their disposal. The locally directed process to date has also taught more granular lessons: strength in an expanding area cannot happen without growth; it requires an institutional home that nurtures and supports; and connections and coordination, desired as they are, cannot be sustained long-term without a core unit.

With the information at our disposal, we have concluded that three measures are needed. In developing these recommendations we have attended to both the needs of data science and the needs of the campus at large. We outline these measures here and expand upon their rationales in later sections of this report:

1. **Organizational form**: Move to create a flexible, innovative academic core of independent decanal stature, a School centered on computation and data science, with a mandate to develop a robust culture of engagement and strong mechanisms of connection campus-wide. Departments, programs, and institutes can use regular faculty governance processes to populate the School initially, fitting into existing mechanisms defined by the Academic Senate and administration and respecting individual faculty autonomy throughout. Faculty in several key units are ready to consider this action, including Statistics, Electrical Engineering and Computer Sciences, and the School of Information, and and our strong sense is that with faculty from these units and possibly others a world-leading School could be built. Faculty not yet organized into programs also need ways to participate and, as appropriate, join, and internal boundaries within the School among departments and programs are open to being reconfigured if faculty wish. Matching the Berkeley model of integrated excellence from undergraduate through graduate education and research, an academic unit of this character—marked by novelty, ambition, supra-departmental scale, and internal heterogeneity—needs focused direction that can pursue long-term priorities with decanal tools. Within this core unit the Berkeley Institute for Data Science could appropriately flourish as a crossroads for application-oriented collaborations, open/reproducible practices, and targeted service to science domains.

   We recommend that the goal be to form the School administratively as a Division of two existing Colleges (Engineering and Letters & Science), although it could also exist as a free-standing academic unit. Despite the unfamiliarity of a cross- or bi-college division, careful examination and consultation with experts have persuaded us that it is administratively feasible, has some noteworthy advantages, and aligns with good governance practice. Where differences of opinion exist among us, they lie in the likelihood or desirability of a trajectory that might move from a cross-/bi-college form to a free-standing School at some point.

2. **Strategic academic plan for faculty FTE**: Invest in an influx of data science faculty positions (over ten years, on the order of 20-25 FTE strength). This will require a substantial number of new philanthropically endowed (or otherwise externally funded) FTE as well as standard faculty lines. Faculty positions are key to both expanding core domains and building broad strength as this area surges. In line with that understanding, a balance of FTE will need to be worked out between the significant number of faculty positions essential to fill out the new School and those in, across, or between other disciplines and units invested in data science. To get the benefit of the influx, we see an urgent need for hard-headed, collaborative development of principle-based allocation mechanisms to provide appropriate incentives for securing philanthropically funded faculty positions and to steward the overall process to an effective end.

   Within this process we advocate for providing a path for faculty across departments to identify targeted application areas for decisive investment of FTE, such as computational social science, biology, and environment. It is essential that where joint appointments are used, they should be constructed using best practices that provide a strong foundation for both individual faculty careers and programmatic success.

3. **Fundraising**: Move data science rapidly forward as a central pillar of fundraising across Berkeley, including new endowed faculty FTEs inside and outside the core unit, support for key programs and institutes, and the new School as a whole. These elements should be coordinated in an overall plan; securing them need not happen all at once for the overall process to succeed.

   Experience at Berkeley and elsewhere strongly suggests that this is a powerful philanthropic avenue to pursue. The robust participation of Berkeley's computing and data science faculty in the industry ecosystem gives us much leverage, creating opportunities to invigorate fundraising practice university-wide. As part of the campus's new strategies around fundraising, this pillar can be constructed in a coordinated fashion that bridges across academic units. We see major benefits for other academic units seeking to raise funds in data science-enabled application areas to be able to align with this university-wide pillar.

To maintain momentum, these three actions should begin without delay. The process to follow will need to coordinate thoughtfully among them but need not for that reason take undue time. All of the constituent elements are necessary and need to be taken together; there should be no half-measures. If there is potential for controversy, we believe that Berkeley can work through it, guided by the interests of the campus as a whole. These measures will require responsible leadership and strategic cooperation among faculty, administration, and Academic Senate in well-defined processes of shared governance. This is an opportunity Berkeley must seize even at a time of fiscal constraints. Indeed, we can see the efforts within this plan as contributing essentially to strengthening Berkeley's strategic and financial position as well as its capacity to plan and execute.

Next steps

We urge that the campus embark on the next steps with all speed. Specifically, we recommend forming a Data Science Initiative (or some such name) as a transitional vehicle to operate for roughly 24 months, under the leadership of a faculty director charged and empowered to move forward with these measures, responsive to the faculty, the campus administration, and the Academic Senate.

No report, no unit, no program, no priority can engineer an academic outcome at Berkeley. What the campus can do is to make a line of development workable, supporting and enabling the faculty who can carry it forward. Ultimately, good things at Berkeley happen institutionally when they can line up with directions that faculty are moving. This report seeks to help draw out the will of the faculty and set in motion organic processes of change, built on shared governance and our common commitment to the university's success.

Reflections

Our committee's remit is data science, and we interpret this phrase broadly, but we also want to emphasize that the phrase "data science" may not capture all aspects of the emerging intellectual landscape we view as justifying these steps. There are broad issues to be found in the interactions among humans, computing, and the transformation of knowledge that are not about data per se. Nonetheless, interpreted broadly, data science indicates a path forward into crucial, societal-level issues that will, in our view, continue to develop and expand for decades. Indeed, in addition to fundamental questions involving data infrastructures, data collection, data analysis, and data-based decision-making, there are a range of other issues including the role of human oversight in the deployment of autonomous systems, problems of privacy, ownership, and ethics in a computationally enabled world, and timeless questions around the nature of human thought and intelligence. Placing these challenges at the center of campus life will be energizing, will be inspiring to faculty, students, alumni, and donors, and will allow us to create and support the faculty strength that will solidify Berkeley's position as a leader of an emerging intellectual and societal transformation.

Part II

6. Organizational Design

Academic organizational design aims to create a vibrant, integrative intellectual community that delivers on research, education, and service. The expectations placed on that community include powerfully advancing the frontiers of knowledge across the fields within its scope and delivering a strong set of education programs that give access to their substance and prepare students for career paths and research. Within the university, we aim to take advantage of our mutual strengths, so we ask that an academic organization also enable and amplify advances across fields beyond it, and (where this is called for) provide a broad educational pillar upon which fields beyond it can build. We believe that by shaping an organization in alignment with these expectations and developing its distinctive sense of societal mission, the public good can best be served.

Organizational design is a key part of reaching our goals for data science—faculty excellence at the heart of the field and beyond it, outstanding education at scale, strategic organizational capacities for the data science area, and realization of Berkeley's distinctive vision for the field. At the same time, organizational design for data science happens in the context of the university at large. We cannot optimize along every axis simultaneously, and sometimes we have to take on local costs to get overall benefits. On the other hand, the strongest solution for Berkeley is one that takes account of the real interests of diverse parts of campus. It is in exploring the different value propositions that robust solutions can be found.

Recommendation

Move to create a flexible, innovative academic core of independent decanal stature, a School centered on computation and data science, with a mandate to develop a robust culture of engagement and strong mechanisms of connection campus-wide. Departments, programs, and institutes can use regular faculty governance processes to populate the School initially, fitting into existing mechanisms defined by the Academic Senate and administration and respecting individual faculty autonomy throughout. Faculty in several key units are ready to consider this action, including Statistics, Electrical Engineering and Computer Sciences, and the School of Information, and and our strong sense is that with faculty from these units and possibly others a world-leading School could be built. Faculty not yet organized into programs also need ways to participate and, as appropriate, join, and internal boundaries within the School among departments and programs are open to being reconfigured if faculty wish. Matching the Berkeley model of integrated excellence from undergraduate through graduate education and research, an academic unit of this character—marked by novelty, ambition, supra-departmental scale, and internal heterogeneity—needs focused direction that can pursue long-term priorities with decanal tools. Within this core unit the Berkeley Institute for Data Science could appropriately flourish as a crossroads for application-oriented collaborations, open/reproducible practices, and targeted service to science domains.

We recommend that the goal be to form the School administratively as a Division of two existing Colleges (Engineering and Letters & Science), although it could also exist as a free-standing academic unit. Despite the unfamiliarity of a cross- or bi-college division, careful examination and consultation with experts have persuaded us that it is administratively feasible, has some noteworthy advantages, and aligns with good governance practice. Where differences of opinion exist among us, they lie in the likelihood or desirability of a trajectory that might move from a cross-/bi-college form to a free-standing School at some point.

Core and connection

In our organizational thinking, a strong core is a given. Without the ability to retain, attract, and support the very best faculty at the center of data science, the rest of our goals for Berkeley are unrealizable. The organizational structure that we form needs to support their academic excellence and to sustain the community and culture that undergird that excellence. Whatever form the structure takes, it has to provide strategic capacity to spur, shape, and adapt what we do in this domain and to enable the significant growth we expect it will see. Strategic capacity involves exercising academic leadership, stewarding resources, and fundraising; it also involves deliberately curating a culture and shaping a vision. Our underlying sense is that the core should be designed internally for integration rather than differentiation. We want to allow multiple specialties to find their home in it *and* make it easier for them to work with each other, ultimately creating space for new combinations and disciplines to emerge.

The rest of this section of our report lays out organizational design considerations for the data science core; the section that follows takes up connectivity across its bounds. Both core and connectivity are necessary and belong together, but a basic observation leads us to take up the core first: there can be a strong core without strong connections, but there cannot be strong connections without a strong core.

The overall strategy is worth underlining: Our ambition is clearly to do both core and connections, embodying a "both-and" strategy with both elements designed well and effectively linked. Connectivity reaching outward combined with strength at the core have been the animating principle of Berkeley's undergraduate data science curriculum. It is not an accident that data science pulls in faculty for whom the prospect of broad connectivity is a powerful attractor; after all, it is a domain that draws strength from its use. But, equally, the web of cross-connections on its own does not do the job. It does not give us long-range tools to secure faculty excellence at the heart of data science, a critical factor at this moment when we face faculty losses just as this field is about to take off. It does not keep moving our outstanding educational programs forward.[25] Nor does strategic leadership capacity get traction without an institutional anchor. Without strength at the core that commits to drive toward these goals and represents them in the decision-making processes of the university, each of them ends up floating and fragile. That outcome is not in the interest of the university as a whole. In our "both-and" strategy, the core is something we cannot let fail. Connections are something we can powerfully invest in and encourage to flourish in the presence of the core.

Form and scale of an integrative core

At least a dozen implementations of an integrative core, drawing together the distinctive academic elements of this field, have been floated since data science at Berkeley became an item of discussion. These include an organized research unit, an augmented graduate group, a next-generation New Initiative Center, and more (listed in an appendix to this report) At the end of the day, some instantiation of each of these forms may end up being used to realize particular parts of the overall vision, but none of them does the work of a core.

In data science we have departments that lead the world in the deep dimension of the field whose intellectual trajectories are coming closer together. We have a collection of outstanding programs whose future evolution, depending on how they are integrated, could help realize the broad and rich dimensions of our vision. We have inviting possibilities for reconfiguring the relationships among these units, which become more real once they are placed in a common organizational setting. We have a set of undergraduate and graduate programs among which

[25] The data science education program is lodged in a temporary home under the Undergraduate Dean of the College of Letters and Science. If it stays there, the task of leading it would be equivalent to building one of the largest curricular programs at Berkeley in a dynamic, research-intensive domain without a closely associated graduate program or an anchor in its research component, at the same time taking on serious negotiation and coordination with units across campus. No ladder-rank faculty member has been identified who would be willing to lead it in this form.

we find leading disciplinary programs, diverse graduate and professional offerings, and a transformative multidisciplinary curriculum growing at the undergraduate level.

At Berkeley we take it as given that our strength draws from the intrinsic continuity from our undergraduate to our graduate education to our research. Historically that is how our university has succeeded in the domains we are strongest at, and it is how we deliver the greatest value to our students and the world. It seems intuitive to think that it should be the same for data science—that anchoring a compelling undergraduate program, for instance, in an academic unit with the faculty who deliver the core of its teaching, educate graduate students, and break ground along the research frontier is a natural move. That is sensible organizational practice in terms of aligning incentives and responsibility. It also gets at something deep about the world: all of these elements of our practice as a world-leading university actually *are* integrated, and the strongest academic organizational design is responsive to this fundamental condition.

We conclude that this core needs to be an academic unit, traditionally formed in some ways, possibly untraditionally in others. Before we talk concretely about organizational capacities, construction, and structure, we should get the scale of what we are talking about into view.

If we look at areas where Berkeley currently has faculty doing research and teaching in the deep and rich dimensions of data science, we find them concentrated in particular departments and programs. Even as their departments encompass much more than data science, those faculty today comprise a significant part of computer science, statistics, and information, as well significant portions of electrical engineering. The group likely brings in elements of computational biology, cognitive science, new media, and emerging aspects of quantitative social sciences, optimization, and possibly other areas that are sprinkled across existing units. The boundaries between these fields have become blurred, with numerous joint appointments pointing to overlaps among departments and programs, despite being separated in distinct colleges, schools, and initiatives. As a body, this group of faculty delivers multiple very large undergraduate majors, operating in two colleges simultaneously (Engineering and Letters & Science). Judging by student trends, it de facto provides an educational foundation campus-wide.

Numerically, the body of faculty in these departments and programs is in excess of 100 faculty today. This core is of the scale of our current Letters and Science divisions, colleges, and largest schools.[26]

[26] The breakdown of academic FTE incumbents in decanal units (for HWNI, quasi-decanal) in 2015-16 is drawn from data maintained by the Campus Budget Office and the Vice Provost for the Faculty and follows the Budget Office's specifications about Agronomist series appointments, non-state-funded FTE, and teaching faculty (all included). Hiring by the four New Initiative Centers and the Haas Institute for a Fair and Inclusive Society is captured in the departments holding their FTE.

Incumbent FTE by Decanal Unit, 2015-16 (ordered by scale)			
Colleges and Divisions			
L&S Social Sciences Division	269.45	L&S Biological Sciences Division	105.95
L&S Arts & Humanities Division	241.00	Chemistry (College of)	58.50
Engineering (College of)	226.50	Environmental Design (College of)	52.50
L&S Math & Physical Sciences Division	148.25	L&S Undergraduate Division	3.00
Natural Resources (College of)	110.85		
Schools			
Business	82.00	Social Welfare	16.00
Law	61.25	Optometry	15.00
Public Health	46.75	Information	12.00
Education	28.25	Journalism	9.00
Public Policy	18.50		
Other			
Helen Wills Neuroscience Institute	6.50		

Decanal character of the core unit

Whatever academic unit we build in this area needs capacity for strategy, operations, and shared culture. To carry that off takes an investment in leadership. We are trying to do something more than combine masses of faculty and programs, or even remove organizational obstacles that get in their way. An academic unit for data science—marked by novelty, ambition, supra-departmental scale, internal heterogeneity, and collaborative integration—needs more than coordination of pieces whose primary reporting lines lie outside. It needs focused direction that can pursue long-term priorities with decanal tools.

Strategic capacity

This decanal unit needs to be built in a way that aligns resources with responsibilities so that it can effectively steward shared priorities and execute on its plans. Further, its leadership needs to represent an integrative intellectual agenda and have the tools to help its faculty shape it. It needs to have the capacity to cultivate a character that is open to application and real-world impact among its criteria of merit when it shepherds faculty careers. It needs to take in constructive work and collaboration across disciplinary boundaries as cultural values, and it needs to build on a shared commitment already evident in many faculty in this area to outstanding education delivered at scale. These values are not foreign to Berkeley, but they are not uniformly represented across it, either, and anchoring them in the organization will take leadership. That leadership needs to be able to exercise decanal capabilities in academic personnel planning, resource allocation, and curriculum planning. In these functions it cannot be a subordinate part of some other decanal hierarchy. It needs independent decanal strength, moreover, to have a voice in shaping the campus's direction and to pursue systematic fundraising.

Practical operations

With decanal capacities, this organization can steward the functions that support faculty careers and research. That is perhaps the critical component that faculty expect from the overarching academic unit to which they belong. The organization can develop external relationships that integrate and project the strengths of its constituent pieces, both in fundraising and in industry outreach. It can participate effectively in developing strong revenue-generating opportunities that combine the strengths of the faculty and units for which it is responsible.

Critically, a decanal unit can administer the full range of undergraduate and graduate academic programs as they currently exist or may evolve:

- **Existing majors and minors** hosted by departments that may elect to move into this unit would be administratively accommodated. In any likely configuration of departments, these include both B.A. and B.S. degrees. Some of these existing programs, as we have noted, are very large.
- **Developing our data science education offerings** at the undergraduate level is significantly easier when there is a dean to look out for their needs, keep them in accordance with graduate-level teaching, and invest the effort in building connections that pay off. Based on trends in existing major programs with affinities to data science, enrollments in the Data 8 Foundations course, and surveys of students, we anticipate that data science major and minor programs may come to constitute one of the largest instructional programs at Berkeley.[27]
- **There are different ways to approach admission of undergraduates**, depending on what type of decanal unit is formed. There is no intrinsic reason to admit undergraduates to this academic unit as freshmen, and no intrinsic reason not to. Both approaches have been workable for Berkeley's colleges and schools. For transfer students the pathways are likewise clear.

---

[27] It would be premature to give estimates of major and minor sizes until such programs are outlined to the point that they can be considered by the Academic Senate. Based on Spring 2016 Data 8 students' responses to survey questions, at least 400 majors and 600 minors annually might be projected within 3 years. At this stage it is sensible to plan for an initial envelope that includes majors and minors 1-2x that size, adjusting as the program is built. Some part of this capacity may be accommodated as reduced numbers of majors in other programs; another part will likely be additive, as students take on multiple majors.

- **The unit can strategically fill out the campus-wide portfolio of professional masters programs**, in coordination with other units who will also want to offer data science-inflected degrees. It can also expand into other kinds of revenue-generating educational offerings beyond the professional masters.
- **For PhD-level education**, there may be value in having a decanal unit's capacities, recognizing that PhD emphases are typically more specialized and distributed in character. Berkeley has relatively successful mechanisms for both disciplinary and interdisciplinary graduate programs that could be hosted by this unit. At this point it is yet to be seen whether a PhD program in data science should be formed, or how other graduate student needs might be better met in relation to domains that use data science, for instance, in the graduate minor (or Designated Emphasis) space.

Culture and leadership

The overall aim of the core unit is integration, not differentiation, given its distinctive and heterogeneous parts. Because of the different trajectories and practices of the faculty and programs that will constitute it, its culture will necessarily involve recognition of and respect for those differences. Both disciplinarity and interdisciplinarity will need room to flourish, at the very least between its constituent parts, and ideally across them and within them. A shared basis for culture will then need to be found in broadly shared principles. Alongside theoretical contributions and conceptual breadth, these will need to include outward-facing openness to real-world impact and success in application. Collaborative contributions will need to be acknowledged in creating an integrative, border-crossing community. While this culture will require thoughtful stewardship of faculty promotion criteria, it is not impossible. The experience of our application-oriented colleges and professional schools will be important to draw upon. It is also important to put a strong emphasis on adaptability and continuous evolution. It is fortunate that the core unit would be able to build on cultural elements of openness, external collaboration, and experimentation that are already deeply anchored in several of the departments and programs that may wish to join.

The leadership of this unit must deliberately steward its culture no less than its academic excellence, as these are overtly linked in multiple ways. Such stewardship plays out in many spheres, including organizational values, academic promotion processes, and strategic planning for the future. It goes without saying that these cultural expectations about collaborative openness, constructive merit, and exploratory interdisciplinary engagement are essential criteria for the selection of its dean.

Internal structure

The internal structure of this organization will need to be defined by the faculty who are part of it. **If we follow the lead of the faculty, the structure will follow.** We would assume the structure can be flexible and adaptive, so that decisions are not once-and-for-all. Given the historical affiliations and heterogeneity of faculty who may participate, we would expect that to get started, where it suits them, they may choose to keep a departmental structure close to what they currently have, then give themselves ways to redefine the boundaries as they shape a collective sense of their needs.

Discussion of options has already been broached to differing degrees inside several academic units (Statistics, School of Information, and Electrical Engineering and Computer Sciences in both its EE and CS Divisions). These conversations are a basis for further discussion but do not predetermine the outcome. If the group of faculty currently housed in a department (one of the above or another) wishes to engage in the new organization as a unit, they will presumably consider options in line with their regular governance procedures. There is no obligation that a department move as a whole. Constituent groups or individual faculty or groups may make their own decisions about seeking a home for their FTE. Other faculty may also join in, individually or in groups. The organization would provide a place to re-house existing centers, NICs, and programs if their faculty choose to join. Exploring departmentalization options could be part of the process for these groups.[28]

Varied possibilities then open up for reconfiguring internal structures. This could happen at multiple points along the organization's trajectory, from its initial constitution to its evolution a decade or more out. Here again, if we follow the lead of the faculty, the structure will follow.

- **There may be interest in forming a new academic subunit** of some form (a department, graduate group, or some other structure) **around aspects of computation and data bridging strongly to fields outside the unit**. Just as a placeholder, we call this "Computational X" (without any presumptions at all about what it might include) and include its exploration as part of the process of defining the new organizational form. (Some additional thoughts toward Computational X can be found in later sections of this report, on organizational design for connectivity and strategic foci in faculty FTE hiring.)
- **New internal configurations may emerge** around emphases that show up in different places in the initial set of departments, programs, and faculty. One opportunity could be the human connections of computing and data. Other opportunities might be found in robotics or in analytical methods for computational science and engineering. These configurations (again, a department, a graduate group, or some other structure) could also attract significant participation from individual faculty whose home units would still be outside.
- **Nucleating new specialties** within the organization or with additional bridges to outside it should be part of its goal. Doing so successfully requires a strategic eye on developments across the heterogeneous fields represented inside the unit, as well as broad attention to changes in the larger disciplinary landscape and the world outside.

Our assumption is that it is more likely that promising reconfigurations can happen if the faculty are brought together in an integrative core unit than if they are held apart.

---

[28] This move would, we expect, be attractive to some groups of faculty who are invested in academic programs in the penumbra of computing especially, and feel they have been hampered by the lack of support that comes from having a strong organizational form or a dean.

Data science is not the only confluence among these faculty and entities; others involve humanist-computing connections, design, and optimization, and more may be forming. Nor is data science contained entirely within these entities; it reaches beyond. Each of the areas involved has deep, well-established subfields that are clearly distinct from data science. The organizational unit needs to accommodate this complexity and allow it to continue to unfold as part of its collective identity.

Options for a decanal unit

There are different ways to realize an academic unit with decanal leadership. Some of them we do not recommend. Conspicuously, a matrixed organization, such as a “faculty” responsible for the undergraduate educational program (with or without elements of graduate or research emphasis), will not be institutionally capable of delivering on our goals.[29] Of the more plausible forms:

- **A free-standing school or college** is a standard option with some real advantages. Outside of data science, a substantial number of colleges and schools exist at Berkeley already. If one were to emerge around computation and data, it would be an immediate peer of the ones that currently exist in terms of scale, academic excellence, leadership capacity, and revenue-generating power. The last of these includes philanthropic opportunities at considerable scale. Compared to a college, a school would be the more straightforward path. **We would not rule out this option either as an end state or as an evolutionary path.**[30]

- **A division within a college** is a decanal unit, as exemplified by the five divisions of Letters and Science. Within L&S, the divisions are distinct, peer decanal entities. Each dean reports directly to the Executive Vice Chancellor and Provost. None is subordinate to another, and the dean serving as Executive Dean of the college may come from any division. Key resources, including FTE, annual budget, TAS, faculty salary savings, and overhead return, are allocated directly to divisions. Promotions, strategic plans, and budget requests roll up separately by division.[31] The divisional option is available to any

[29] If the organization’s faculty and programs were to remain in their existing organizational settings, being matrixed across for particular purposes and headed for those functions by someone with the title of dean, strategic planning and resource allocation capacity for data science would be compromised by dependency on academic units whose positioning has so far not supported its emergence as a priority. This is a structural observation that gains strength from observations outside of data science. While there are attractions to cross-cutting organizations, strategic shortfall has been the experience of quasi-decanal units at Berkeley and elsewhere, especially when they aim to provide a strong core without having a peer relationship to academic deans in academic planning, resource allocation, or fundraising.

[30] The choice between school or college would depend partly on curricular and programmatic intentions (separate undergraduate admissions procedures, particular options for degree programs), partly on administrative considerations like approvals and transition planning.

[31] Departments are associated with divisions, not generically with the college. Requests and allocations go directly from divisions to the campus level and back; there is no overall college intermediary to report through. Divisions are knitted together through college governance structures and the ad hoc relationships that come from faculty and student activities. The college provides an academic envelope

college at Berkeley; only L&S has exercised it. **For data science we believe that a division located inside any one of our existing colleges is unlikely to attract broad faculty support, so while we list it here, we do not recommend it.**[32]

- **A division within *two* colleges that is denominated a school.** We explore this non-traditional option in the subsection that follows. **This is the solution we recommend pursuing if the faculty agree.**

A school as a division within two colleges

The innovation we suggest comes from taking seriously the outward focus of this unit and the sense of anchoring that many faculty feel in the organizational homes that have supported them until now: the College of Engineering, the College of Letters & Science, and the community of professional schools. We observe that even as faculty may seek to develop autonomous strategic capacity for data science and grow a new academic core unit, they may simultaneously hope to find ways to remain philosophically and administratively connected to those hosts—to retain strong, long-standing connections, prevent competing duplications of effort, retain and extend reputation and standing, harmonize strategic directions, develop revenue and fundraising avenues, utilize common foundations and functions, and then build on those linkages to strengthen their outward-facing connections to many departments and programs.

We thus see a solution in an organizational opportunity that we ultimately recommend pursuing if the faculty agree: first, that a single new decanal division for computation and data be formed, that it be housed in *both* of the Colleges of Engineering and Letters & Science, and that to signal its distinctiveness and engagement with the world, as well as its commitment to professional education, it be called out as a School if the formalities allow.

While the reader's mind may naturally move first of all to potential administrative complexities, we describe this organizational option first from the perspective of faculty experience. Faculty would have positions within this organization as in any other school, division, or undivided college. The arrangement includes the possibility of joint positions with units anywhere inside or outside the two hosting colleges, as is the case today. The organization would have internal departments that provide research and teaching programs as usual, governed in their usual

---

responsible for admissions, granting of degrees, advising, and coordination of fundraising, and it provides some shared infrastructure for business and other operations.

[32] That is true whether it is a reframed existing division (for instance, the Division of Mathematical and Physical Sciences in L&S) or a new division (for example, inside the College of Engineering). The reasons are many, detailed, and in our observation powerfully felt. They stem from the expectation that no college we currently have at Berkeley would be able to provide the package of cultural, operational, and strategic elements that would allow this new boundary-crossing field to thrive, particularly compared to a free-standing school. Opting for one of the existing colleges would also be seen as sending a signal about the organization's orientation, and possibly its ownership, in a way that would limit the paths of its graduates and the openness of faculty across campus to engage in partnering with it.

way.[33] It would leverage some of the organizational capacities of large existing colleges, which admit students into degree programs and confer degrees. Existing degree programs could thus be preserved. While the cross-disciplinary nature of computing, inference, and data calls for an unusual degree of connectedness in organizational structure, faculty are not matrixed across decanal units or required to be split between distinct constituencies.[34]

From the point of view of administration, governance, and leadership:

- **The construction has administrative feasibility.** Because of the independent status of each decanal unit at Berkeley, the dean of such a unit has the same capacities as the dean of a school, just one that participates in two colleges instead of zero or one. In the Berkeley construction of a decanal division, being part of two colleges means two things: participating in two colleges' governance structures and drawing from two colleges' administrative and academic support (such as admissions and advising). On the side of governance, the dean of this unit would sit as a peer in both colleges' governance arrangements, as well as presumably in groups involving deans of professional schools. On the side of infrastructure, with clear lines of reporting and memoranda of understanding it is administratively and academically feasible both to bring existing programs into this structure and to constitute new ones.[35]

- **The leadership model is workable.** There is only one dean of the unit, drawn from the faculty of its constituent departments and programs. Like every other dean, this dean would report directly to the EVCP. He or she would receive allocations of resources, take responsibility for stewarding them, set strategic direction, and execute on it. Deans within L&S do this already, at the same time as they partner with their peers in the College. The key consideration is that each dean is independently equipped with decanal tools, whether he or she sits in zero colleges, in one, or in two.

- **Fundraising would require both the usual decanal capacities and cooperation with other units.** Like all other decanal divisions, this decanal unit would have dedicated fundraising capabilities. It would also need to work cooperatively with its counterparts in the colleges and the university. It is important to say that in whatever way this decanal

[33] Despite issues of nomenclature, there is no intrinsic complexity to having divisions of a department inside a decanal division. Those are separate levels of organization and do not cause confusion in practice. As an example, Molecular and Cell Biology is a department in the Division of Biological Sciences. It contains five internal Divisions.

[34] Faculty in the school would be in the position to make a collective decision about how individuals and programs would describe their relation to the two colleges. One could imagine all the faculty and programs identifying with both colleges. Alternatively, each could identify with the college(s) in which their undergraduate major programs were embedded. There are surely other options; the point is that it is for the body of faculty involved to decide.

[35] This is not as foreign as it may seem. L&S and the College of Engineering already support bi-college pairs of majors (e.g., CoE EECS and L&S CS) with working arrangements around TAS funding, class enrollment priorities, and some shared advising staff. Programs have existed shared between other colleges before, as between CNR and L&S. Though we are not recommending matrixing the faculty or the academic programs, we find matrixing of support functions a conceivable approach.

unit is constructed, it will have to work in coordination with others. Such coordination can be done from the inside (using internal governance mechanisms within the colleges, memoranda of understanding, etc.), from the outside (through campus-level mechanisms or direction from leadership), or both.

- **This option would address other organizational considerations** that arise together with the emergence of data science. One consideration the FAB has heard is the value of not sending signals of separation and siloing. Although we find this argument quite malleable (free-standing schools and colleges at Berkeley are not intrinsically silos, indeed, can be deeply collaborative with other units), we mention it as a perception that we have heard expressed. Another consideration is the value sometimes attached to not presenting a decision about faculty potentially leaving the College of Engineering, where EECS is a strong presence and valuable contributor. Faced with a strategic choice of the magnitude of data science, the strength of existing organizational configurations cannot be the deciding factor. All the same, it is important for campus-level considerations, as it points to real implications, just as the continuity of ongoing degree programs does.

- **Cooperation of the colleges would be called upon** in constituting the school as a division within them, rather than a free-standing unit. All involved deans would need to be accepting and flexible while this works itself out in the first few years. L&S would presumably need to add another division, forming a sixth, and Engineering would presumably explore options such as creating two internal divisions (for computation and data and for the remaining constellation of engineering faculty), possibly with an executive dean of the college as on the L&S model in a collaborative, non-hierarchical governance structure. It would be possible to view the division for computation and data as a kind of “hinge” between these two colleges.

- **It appears feasible to attach the terminology of a School** to a unit whose organizational structure and capacities match University of California system understandings of that term, whether or not it is part of a college.

- **An appropriate level of visibility to philanthropy is secured**, in the FAB’s assessment, by having a dean and an organizational identity as a school. This would be a valuable point to proof with advisors from the philanthropic sphere.

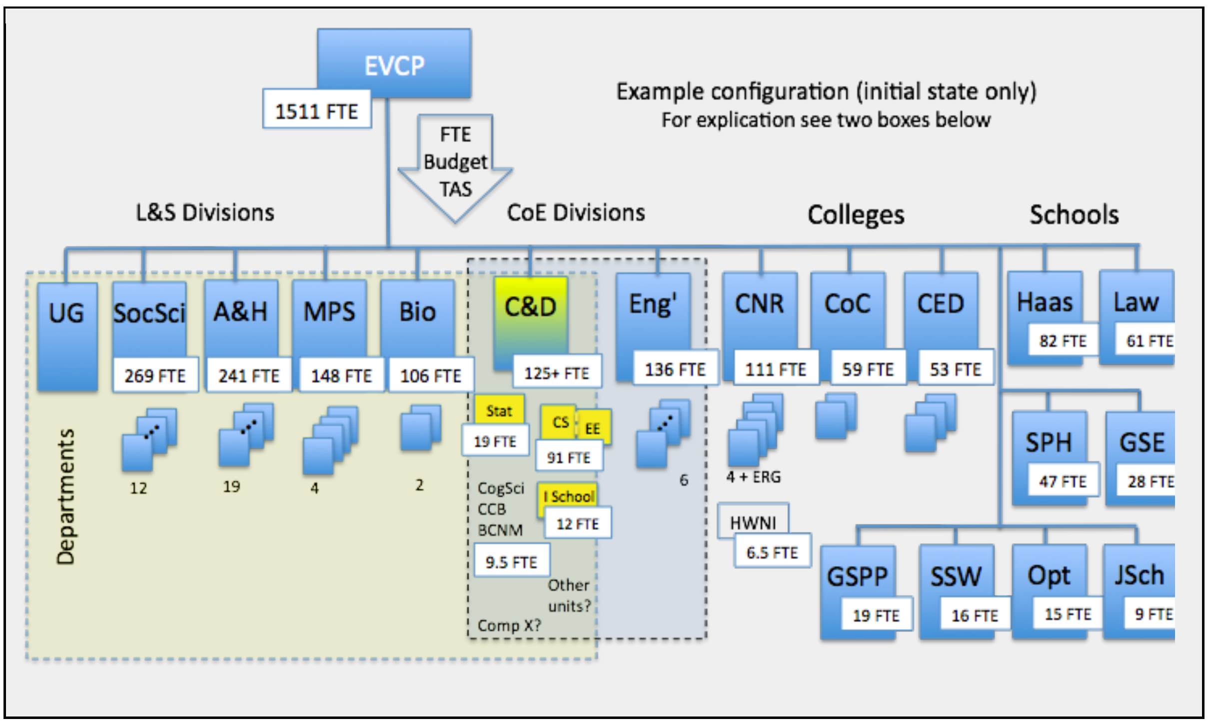
- **If it proves useful to evolve to a free-standing structure**, the initial groundwork will have been laid.

It will be for the faculty and the campus to decide whether this “cross-college” or “bi-college” path should be pursued. The FAB has spent considerable time reviewing it and finds it has attractions that make it our primary recommendation. All the same, we see that heading down this path would entail overcoming uncertainty and establishing new practices. We devote additional space to these considerations in our section on situational challenges below.

Possible organizational configurations

For the sake of concreteness, the FAB has found it useful to work with rough sketches of the varied configurations such a school might take. This is all modeling in advance of an actual process of constitution. Ultimately, the decision to form a school of any particular configuration comes at the end of a process that is grounded in faculty self-organization and Senate and administrative review.

All advance discussion is thus abstract in the strong sense of being removed from reality. It is only after deciding to form a school that it is really possible to bring to the fore the deeper issues about how it should constituted. All the same, to make that sense of reality come closer, we offer a sketch of one among many possible initial configurations, then talk about possible evolutionary outcomes.



Example sketch of initial configuration

- This is not the necessary outcome, just an illustration. Minute details of placement of lines and boxes do not imply presumptions about units inside or outside of boundaries.
- This is an org chart representation, which illustrates reporting and resource allocation boundaries. It obscures the fluidity and relationships inside and across boundaries that actually characterize academic life.
- We assume departments cannot straddle college boundaries. It is a question for college and faculty governance whether departments can straddle decanal boundaries inside a college by means of internal mechanisms that respect resource allocation and reporting relations.
- We give current names and sizes for existing units, for simplicity assuming all faculty choose to move with them. It should be assumed these may evolve over time.

- For a free-standing school, the one change to the diagram is to move the decanal unit for computation and data outside the boundaries of Letters & Science and Engineering.

Multiple evolutionary outcomes are possible from any initial configuration

- From any initial configuration, internal boundaries may be reconstituted as the unit evolves. The standard processes for faculty-led change applies (departmental, administrative, and Senate governance, as per the Berkeley Review Process Guide).
- This evolution may involve the emergence of clusters of faculty, who may be formed as departments, divisions within departments, graduate groups, programs, etc., with a porous boundary to others outside.
- Some forms of clusters need to stay inside org-chart boundaries; others are deliberately set up to do interdisciplinary work that crosses those boundaries.
- The presence of a computation and data decanal unit may cause new clusters to form inside, across, or outside its boundaries. There is no presumption that all academic activity around computation and data would fit inside it.

Forming the new school: principles and process

Forming the new school will take coordination among the faculty, the Academic Senate, and the administration. The process starts from organic articulation of faculty interest and follows paths in line with current governance mechanisms and a constitutional process. Here we sketch a model that would enable structured development of collective planning, so existing units can engage and individual faculty can exercise options as needed. Once a concrete proposal is prepared by faculty, the model fits into the review and decision-making steps assigned to the Senate and administration at campus and UC levels (Berkeley Review Process Guide, University of California Compendium).[36]

Principles

Starting from shared principles allows orderly steps to define the organization's internal structure and formulate or incorporate new units. In this emergent process, the place to begin is to articulate the core organization's role and mission. We recommend beginning with a self-identifying group of existing units that would consider joining the core, drawing on their existing faculty governance processes. We suggest the following candidate principles as a starting point for faculty dialogue:[37]

---

[36] Berkeley Campus Review Process Guide for Academic Programs and Units (2011), http://opa.berkeley.edu/sites/default/files/ReviewProcessGuide.pdf; Compendium: Universitywide Review Processes for Academic Programs, Academic Units, & Research Units (2014), http://www.ucop.edu/institutional-research-academic-planning/_files/compendium_sept2014.pdf.
[37] For long-range success, the decision to be part of the new organization should be framed as a commitment to its mission, intellectual agenda, and opportunity, rather than short-term considerations of resources or influence.

1. A substantial fraction of the collective academic work of the faculty in the unit has computation, inference, or data analytics as a fundamental element or pertains substantially to areas associated with data science.
2. The nature of that work brings connections with other disciplines.
3. The faculty of the unit are open to the collaborative, open reorientation of vision and culture that the new school would be aiming to realize.
4. They are committed to and prepared to take on the broad educational mandate of the new school in meaningful connection to their own research.

Statistics, Information, and the divisions of EECS (Computer Science and Electrical Engineering) or the department as a whole may find that they match these principles. Computational biology, new media, and cognitive science may also. The opportunity may be interesting for other units as well, for instance, IEOR and Demography, especially if their strategic directions are considered. Such units have well-defined faculty bodies and governance processes for deciding collectively whether to take an institutional action, i.e., to join or not. It is also possible that newly self-constituting groups of faculty find they also match the shared principles, in which case there should also be procedural ways for them to engage.

Process

Assuming that a critical mass of faculty elect to participate, they can form an interim governance structure responsible for producing the core unit proposal required in the BRPG and carrying it through the approval processes. In formulating the proposal, the constituent faculty would have the opportunity to determine specifics of their internal organizational structure, which might reflect their historical boundaries or might involve a refactoring of the faculty body. It would also detail the external connections, disposition of degree programs, and processes by which additional units or individuals may join.[38]

A distinct challenge arises for potential units that have not yet formed to the point that they can have distinct recognition and governance mechanisms. Their formation might take considerable time, in which case the process of integration could be performed by the new core organization once it is established. However, the decision to form the core organization is likely to accelerate that process and bring new potential unit(s) to the fore. The faculty who might choose eventually to participate with part or all of their faculty FTE would need their own avenue for engagement. Thus there also needs to be a structured process of inviting and convening a separate group of potentially interested faculty who are not otherwise included in the process outlined above. This group would need to develop its own criteria for faculty participation and governance. In order to be part of the constitution of the core organization, it would need to be moving forward on the timeline set by processes in existing units; if that timeline is too fast, it will need to be integrated later. The placeholder for this group in our model is “Computational X,” but it could take many forms and would necessarily emerge from faculty dialogue.

[38] Should they choose, individual faculty will have the opportunity to exercise the option not to go along with their units, using standard mechanisms for relocating their faculty FTE. The cognizant dean(s) and the Vice Provost for the Faculty shepherd such processes.

The overall process of formation should be clearly spelled out and placed on a definite timeline. Given extensive discussions to date, the FAB sees every reason to move forward with engaging faculty governance processes. Moreover, we see major costs in delaying.

1. The constitutional process should be started in Fall 2016. It will draw heavily on existing governance mechanisms inside units and require attention to engaging groups of faculty outside.
2. The interim goal should be to have an organizational proposal for the new core unit entering review by the Berkeley Division of the Academic Senate and the campus administration (as per the Berkeley Review Process Guide) by spring 2017.
3. A full transition and operational plan will need to be drawn up.
4. Necessary approval processes can then be engaged, including system-wide levels as needed. The goal is to have approvals in place and transition planning complete so that the new unit can be formed by July 2018.

We offer more detailed scaffolding of this process later in our report. Processes of fundraising, interim administrative structures, alignment of resources, and preparation of academic program changes should proceed in parallel and in coordination with the timeline for establishing the institutional organization.[39]

Conclusion

In sum, the FAB recommends that processes commence to enable a strong core of faculty to be formed into a new school of one of two organizational constructions. This involves creating a decanal entity with a strong independent mission, a seat at the table in shaping and executing university priorities, and the responsibility to deliver a major component of the university mission in research, teaching, and serving the public good. It needs to work closely, as a peer, with other decanal entities from a position of strength deriving from excellence in research and teaching. It needs the ability to fundraise actively and set priorities. It needs to build the administrative structures that allow its department-scale constituents to operate effectively, both in pursuing their independent priorities and in working together as a whole in pursuing their common missions. It needs to provide a context where new intellectual fields can be nucleated and develop strong connections to the rest of the campus. It is to this second aspect, connectivity, that we now turn.

[39] These transitional processes can begin before any new space may be required. The transition will draw on existing administrative infrastructure in departments, colleges, and schools. For some time it will be a surge process, and that will present an extra load for the campus to carry. Additional details are given in the situational challenges section.

7. Connectivity

Building and sustaining connections between academic disciplines is hard, in data science just as in other transdisciplinary domains. Connections take hold only if they brings benefits to both sides. The FAB thinks there are huge opportunities in connecting across the constituencies of data science, and we believe these connections will become even more important in the future than they are now. Creating thoughtful, realistic ways to support them will be a major investment in Berkeley's leading role in the field, so we think of connectivity as a key part of our organizational design.

With the rapid spread of the tools of computing and data analysis across the university, faculty and students in every domain are engaging in data-intensive work. Several professional schools are also launching, or considering, data science-related degrees. This diverse landscape of data-related research and education is a source of strength for the Berkeley campus. In our core-and-connections model, the new school would not expect to be involved with, and even less to own or control, these activities. Instead, as a campus-wide educational resource and as a source of research expertise and potential partnerships, the school would recognize that data science activity will continue beyond its boundaries at the same time as it flourishes internally. Ideally the openness and connectivity of the school will support mutual awareness, help avoid the wasteful duplication of efforts, and over time create an environment for more cross-campus collaboration. Above all, there is no way to engineer collaborations and connections with Berkeley-caliber faculty. What we can do is experiment with approaches and mechanisms that make these easy to find and rewarding to do.

Data science as a field of connections

Data science crosses intellectual boundaries with a huge array of capacities to address problems in different domains, with similar tools often applicable across diverse fields.

The potential for collaboration and integration is reciprocal:

- There is a clear benefit to applying existing and future intellectual developments in data science (ideas and methods, algorithms and software) to domain areas.
- As new statistical and machine learning methods and algorithms have shown, cutting-edge developments in data science can quickly become useful to a broad spectrum of application-domain researchers.
- Domain areas where data science is used can feed new problems to foundational researchers to work on.
- Domain areas are where distinct expertise can be found on challenges of modeling or research design, limitations of available data sources, and critical embedded issues of the ethical and sociopolitical entanglements of data science research.

Platforms for collaboration and integration connect a disciplinary core with other disciplines and vice versa, adding to the strength of the respective fields.

- Economies can be made in both human and financial resources, since sharing of knowledge means similar problems need not be solved again and again independently.
- Coordination and streamlining allows identification of areas where expertise is needed.
- When similar problems can ground conversations across disciplinary boundaries, shared venues, time, and language become the bottlenecks, rather than intellectual power.
- Network effects come in when researchers and educators in different domains recognize common features to their challenges.
- Networks of connections also create pathways of percolation through the institution, and they have campus-level effects. As faculty talk to faculty, they provide mechanisms for changing the academic landscape over time.
- Uncertainty pervades all settings in which data is used as partial evidence to infer underlying truths, form predictions, and make decisions. The need to make honest assessments of uncertainty, gather observations so as to reduce uncertainty, and communicate uncertainty are part-and-parcel of the modern scientific method.

Data science may be coming into a new position across many disciplines. It may play a fundamental role like that which mathematics has played in science and engineering for a long time. Because of this transition, designing for connections matters especially for data science.

- We may make it possible that competence in data science among our graduates and affiliates (undergraduates, graduate students, postdocs, and visitors) could become a new hallmark of Berkeley. This across-the-board competence could increase the competitive edge and market value of Berkeley graduates in all fields that benefit from data science.
- In data science, designing for connectivity is a way of designing for inter- or trans-disciplinarity, both for individual collaborations and through network effects.

Facilitating connections in data science

How do we design for connectivity in data science? We can see standard factors of success that that we want to build in, either anchoring them organizationally or seeding them culturally.

- **Convening functions**
    - Open seminars (lunches, etc.) that combine great intellectual content, smart participants, and tolerance of multiple perspectives.
    - Forms of shared space that make informal, unplanned encounters possible.
    - Light forms of discretionary resource allocation to make it easy to get something started without lots of permissions and paperwork.
- **Faculty participation across departments and programs**
    - A voice in strategic planning and faculty hiring in more than one department (zero percent appointments can do this; there may be other ways).

    - Joint appointments if it is institutionally workable and attractive.[40]
    - Co-teaching, participation in shared curricula.
- **Graduate students**
    - Students working with multiple faculty advisors. The construction of PhD committees at Berkeley is an incredible asset. This can be fostered.
    - Fluidity in students' departmental paths and allegiances, so faculty can engage with (and recruit) students in more than one program.
- **Academic culture and standards**
    - Openness to respectful discussion and debate from different perspectives.
    - Recognition that publications in other disciplinary settings follows a different set of standards as in the home field.
    - Berkeley's respect for intellectual quality, even if research takes longer to produce outputs, as work across boundaries often does.

Elements to build in

We believe Berkeley should invest in building a strong core to enable strong connections, and we want to enable ways for faculty across campus to have ways to link themselves in. We wish the experience of connecting to be expansive and open-ended, not monopolistic or domineering. That is, we want to create a platform for people engaged with data science across campus to find an anchor in or an overlay over the core unit, as they choose.

An ethos of community, openness, and respect

Parts of the core are already marked by a long-term culture of collaborative engagement across boundaries. Statisticians, most conspicuously, have a long history of joint appointments across the university and an approach to collaboration that involves seeking out shared problems to work on. Building on such foundations, there is an opportunity in data science to form a vibrant and inclusive human community (faculty, graduate students, postdocs, undergraduates, and visitors to campus) that values teamwork and collaboration, making the whole more than the sum of its parts. We hope to shape a community that is ready for open and respectful conversations and discussions on multiple fronts, including resolving conflicts. When participants have an investment in outcomes that depend on this kind of conduct, they are more willing to spend time to approach other points of view, intellectually and personally, and to recognize that optimizing outcomes in collaboration may take more than simply optimizing for individuals' short-range results.

Inclusion is a huge opportunity for Berkeley when we build in this ethos of community, openness, and respect. Respect for diversity of experience and perspective in data science should be common sense. We have already developed programs that carry this spirit in key departments, in D-Lab and BIDS, and in the undergraduate data science curriculum. Our intent

[40] The workability of joint appointments depends on career stage, culture within departments, and good models for academic advancement (e.g., tenure processes that are an OR, not an AND); there is no one-size-fits-all solution we see.

is to develop data science at Berkeley in a way that can break down barriers of exclusivity and do justice to our public mission. When our university takes on data science in a way that is respectful of diversity, equity, and inclusion, it can change the face of the field.

Along with horizontal forms of peer-to-peer (e.g., faculty-to-faculty) connection and supportive elements of advising relationships, we can also create ways to inclusively engage different segments of the academic community.

- **Undergraduates** can catalyze vertical integration between academic strata when they move from sitting in classrooms to working in labs and in groups.
- **Graduate students** are a key element of the academic community and provide connections between undergraduates and graduates as well as between faculty in disciplinary fields, as we have seen.
- **Postdocs**, with a major focus on forging new research, can provide the glue between different disciplinary domains, bringing them together through developing methods and tools in data science. Postdocs in particular are well-positioned to develop capacities that let them achieve strong outcomes in data science research that crosses boundaries, including interpersonal and communication skills and knowledge of the intellectual substance and terminologies from multiple fields.

Platforms that serve and draw people in

Platforms and avenues of outreach can move people across boundaries for longer or shorter periods of time (a semester of leave, a weekend for a workshop, an afternoon for a speaker's talk pitched to audiences outside the core). These require a space that is perceived as being open to others coming in, as well as some commitment of resources and well-tuned attentiveness to what would be of value to colleagues in other domains. Many of the hub-like interdisciplinary institutes in data science at other universities have lessons to teach us about how to make on-the-ground transdisciplinary real. Experimentation is necessary, domain by domain. The possibilities include:

- Open physical locations with a welcoming neutral environment.
- Coordination building on the undergraduate data science curriculum's core-and-connections structure.
- Workshops that address both data science tools and the methodologies behind them.
- An open seminar series that alternates between foundational and application domains.
- An expanded consulting service.
- Rotations and apprenticeships for postdocs and graduate students.
- Vertically integrated mentoring programs.
- An inclusive annual retreat or symposium.
- Forms of communication such as a blog.
- Ad hoc joint research support (e.g., workshop development) as seed money.
- Semester-long programs on the model of the Simons Institute.
- Opportunities to take a sabbatical inside the core.

For many parts of this package we have a promising platform already in the Berkeley Institute for Data Science. BIDS can provide a crossroads where researchers using data science across the disciplines can come together in a neutral space on no one's home terrain. At critical mass, it can have strong network effects: allowing researchers to recognize shared problems, making space for people to collaborate on them, and brokering connections that can help them get resolved. It is a platform, too, that can be used to deliver targeted service to application domains, with particular strengths in software and data systems, propagation of data science tools and methods, and open and reproducible practices. In the context of Berkeley's data science ambitions, it can be a central place where undergraduates, graduate students, postdocs, and faculty can come together outside of departments and programs.

The FAB feels strongly that BIDS is a valuable asset in the service of data science connectivity and should be deliberately mobilized to that end. We see it fitting as a natural partner to the new school, given their missions, and serving as a welcoming portal, ideally with a long-term plan for sustainability incorporating both grant-funded projects and philanthropic support.[41]

Relationships through formal affiliation and organizational integration

A longer-term form of relationship is created by placing people from connected domains in some way "inside" the core. This only works by choice, domain by domain, and it will necessarily evolve incrementally and experimentally as new lines of research and teaching emerge.

At the individual level:

- Established forms of individual affiliation like zero percent and joint faculty appointments with departments in the core.
- Looser forms of affiliation, such as short-term or long-term "visiting" status, or joining some kind of advisory board.

Groups of faculty in computational or data-intensive areas may seek out formal affiliation or organizational integration with the new school. The FAB urges the university to pursue this route, ultimately guided by what faculty articulate as their need:

- Lighter versions include
    - The faculty body involved in a Designated Emphasis (a graduate minor, for which we already have three related to data science)
    - PhD-offering graduate groups in the space between departments
    - Augmented graduate groups with some core FTE[42]
- As options become more solid, as well as in the case of existing New Initiative Centers, we move into the territory of possible departmentalization.

---

[41] BIDS can be joined by other institutes, new or existing, as it makes sense. That evolutionary process can first be realistically discussed as the next steps unfold. Where faculty interest and resources are available, there may be value in nucleating more focused institutes for distinctive domains or for cross-cutting functions, such as data science in the service of the public good. Institutes can be inside or outside the core unit.

[42] The Energy and Resources Group is the example currently existing on campus, situated in CNR.

In our schema, **"Computational X"** (on the model of computational biology, with no pretensions on any final designation) serves as a placeholder for collective forms of organizational integration up to one or more new departments. The model will have different attractions domain by domain, from robotics to computational science to digital humanities. We expect it will be a point of serious consideration for at least two distinct kinds of faculty, either faculty at Berkeley now or new faculty we may want to recruit:

- **Faculty who feel strong intellectual affinity with colleagues in the core** (sharing similar computationally or methodologically focused research agendas, for instance) and simultaneously have part of their identity in an application domain.
- **Faculty who identify primarily with application-domain disciplines** where they are bridgeheads of high-end data science approaches, if they want to have access to an intellectual environment or promotion practices that may differ from those homes.

If one or more groups of faculty precipitate out in the "Computational X" space and seek to engage through at least partial commitment of their own FTE, they may find it appealing to have the new school as an incubation venue. Its investment in novelty would be their opportunity, its shared resources would be part of their platform, and its dean would be their advocate individually and as a group. In return, they would need to weigh their collective capacity to thrive within the new school and help shape its agenda along with the other programs it will also contain.

Instantiating "Computational X" may be a powerful solution to the creation of new computationally or data-related interdisciplinary specialties, which typically have had a troublesome status at Berkeley. However, "Computational X" is not an exclusive move prescribing that developments must happen inside the core rather than elsewhere. It is a placeholder for a conversation that we think needs to happen. Beyond biology, we see opportunities for computational social science and possibly for the domain of the environment, all subject, of course, to the intrinsic sense of faculty in those domains. We comment on some of those opportunities in a later section of this report.

Conclusion

All of these commitments to create connectivity require resources. They also require supple coordination, an experimental mindset, and administrative support. Without prescribing any decisions that would be taken by the new unit, we think this programmatic portfolio for connectivity would take faculty-level leadership. One way to realize it would to mark out a leadership role in the new core organization, such as an associate dean.

It is an unusual approach to data science to create both a strong core and strong connections, at least in the academic sphere. Both are necessary, and both need to move forward. We think Berkeley has a distinctive constellation of assets that provides a very strong basis to build on, plus a set of cultural preconditions that make us hope for success. The FAB's work has foreshadowed this process that needs to continue and expand. We think it is absolutely essential that the exploration continue, and indeed be ramped up, even as a new school is in the process of formally coming into being.

8. Faculty FTE

The future strength of many of Berkeley's programs will depend on securing exceptional faculty with data science strengths. While one-off mechanisms (individual hires department by department) continue to be an essential part of the strategy, we expect they will not be enough to add strength on the needed timescale or achieve critical mass. This matter is absolutely crucial to address if Berkeley is to be best in the world.

Recommendation

Invest in an influx of data science faculty positions (over ten years, on the order of 20-25 FTE strength). This will require a substantial number of new philanthropically endowed (or otherwise externally funded) FTE as well as standard faculty lines. Faculty positions are key to both expanding core domains and building broad strength as this area surges. In line with that understanding, a balance of FTE will need to be worked out between the significant number of faculty positions essential to fill out the new School and those in, across, or between other disciplines and units invested in data science. To get the benefit of the influx, we see an urgent need for hard-headed, collaborative development of principle-based allocation mechanisms to provide appropriate incentives for securing philanthropically funded faculty positions and to steward the overall process to an effective end.

Within this process we advocate for providing a path for faculty across departments to identify targeted application areas for decisive investment of FTE, such as computational social science, biology, and environment. It is essential that where joint appointments are used, they should be constructed using best practices that provide a strong foundation for both individual faculty careers and programmatic success.

Allocation of faculty FTE at Berkeley is steered by processes of shared governance. This arrangement is a fundamental pillar of our excellence as a university. Between the administration and the Senate, the campus must reach a shared understanding of a strategy for academic planning for data science, and we must collaborate at all levels to carry it out.

It is not for the FAB to prescribe how this process should be implemented. The FAB can, however, help articulate starting points for our approach. We can ask the question abstractly, how might strategic academic planning be done for data science? For several reasons, answering this question does not begin by listing a sequence of research topics with a faculty count in each. If it is true that data-enabled discovery and analytics is changing the nature of inquiry in a very broad range of disciplines, the incorporation of faculty representing these aspects must eventually take place as an integral element of the normal request and allocation process. In the short term, the vast majority of faculty performing research and teaching in data science and related areas will be faculty who are already here. However, those faculty are already fully committed to delivering disciplinary research and teaching today. The basis for determining the FTE investment is that of covering the surge. It needs to address a portion of the new and burgeoning demands for research and teaching in data science so that existing programs maintain strength during the transition until shifts in emphasis can take place and routine hiring (replacement) processes can address the need.

It may seem wonky to build a model, particularly at a moment when many other universities are just raising tens or hundreds of millions of dollars to fund new faculty positions. However, a model is a good basis for sound planning and reasoned discussion. With a model we can identify key choices and assumptions, generate confidence that we are operating with the right size of effect, check whether we can mobilize resources that are adequate to achieve our goals, engage in ongoing assessment of our decisions and be fluidly adaptive regarding them, and lay out what some of the qualitative consequences may be.

Modeling an investment in faculty FTE

We expect that advances in data science will be felt across the university, with faculty engaged at levels ranging from basic comprehension of its findings all the way to advancing its intellectual substance. We are not there yet in sufficient faculty strength for what the future will need, so we want to model the transition required to get us to this new steady state. Assume this is a ten-year goal. Along with financial resources, a university's fundamental resource for research and teaching is its faculty, so we take faculty headcount and faculty FTE strength as the basic units. We get some size of the *overall capacity* that needs to be available in data science at the end of the process, plus some scoping of *change* we need to accomplish and the *surge* to get us there. We also get some sense of the *character and scale* of the necessary approach to FTE planning, which feeds into fundraising as well.

Overall capacity (current faculty plus new faculty)

We start from a ballpark sense of the overall capacity in data science faculty. It need not be very refined. We focus on faculty with some level of substantial expertise, from competent, active use all the way up to specialty. We can choose to set the end state (new steady state) to ten years out. In ten years, say, we imagine that with a well-executed strategy, Berkeley, with its 1500+ faculty, would sensibly have roughly 250 faculty (headcount) across the campus with the ability to engage with data science at this level of expertise. It will show up in academic units all over the campus in their research and their teaching. For modeling purposes, we can take research capacity and instructional capacity to be roughly mirrored. This should hold true at the campus scale, though of course locally there may be deviations. (We have more detailed capacity to estimate on the instructional side because of modeling that has been done based on the blueprint for the data science undergraduate education program, though that exercise will bear updating as the curriculum gets built out.[43])

Headcount is different from FTE strength, as many faculty will not have their full effort in this area. FTE strength in this area measures effort available for it in teaching or, more notionally, in research. Data science activity will be spread across this group of faculty with different levels of intensity. With 250 faculty headcount, or 15% of the campus, having some substantial level of data science in their intellectual portfolio, that could translate to a third of that amount of FTE strength, roughly 85 FTE, or 5% of the campus strength in the end state. (This assumes first-order constant faculty numbers overall at Berkeley.)

Among this group of faculty with substantial expertise, we model them using three categories based on their data science effort and emphasis:

Category X	100% effort	fully in data science (all data science, all the time)	1 FTE = 1 headcount
Category Y	50% effort	split attention between data science and something else (e.g., another specialty, an application domain)	1 FTE = 2 headcount
Category Z	10% effort	have data science expertise (a small part of what they do)	1 FTE = 10 headcount

There can be different solutions for distributing the total headcount and FTE strength for the end state across the three categories with these effort levels. These solutions depend on policy choices about the desired profile of our faculty in data science. We can have a policy, for instance, to aim at having equal amounts of faculty effort (FTE strength) in categories X and Y. Then the outcome would look roughly like this:

Category X	100% effort	35 headcount	35 FTE
Category Y	50% effort	70 headcount	35 FTE
Category Z	10% effort	140 headcount	14 FTE
All involved (total)		255 headcount	84 FTE

Obviously, with different assumptions, the details would be different. Still, it gives us a fair sense

[43] Our sense of the educational needs maps onto our sense of the research engagement (amount and trend) on the campus in relation to data science, but this is a more qualitative assessment.

of the distribution of faculty effort, and the model's assumptions are there to be played with.[44]

To make a change, create a surge

The numbers above are targets for overall capacity ten years out. They include faculty currently at Berkeley and faculty not currently here. It does not seem feasible to achieve this ten-year positioning without some enabling investments of faculty lines. While a major portion of this strategic plan can take place via a natural shifting of faculty interests, department foci, and standard FTE replacements, that process alone will not accomplish the change that reaches the ten-year goals. So we need to design an approach to targeted hiring. This approach should be thought of like surge space (in FTE space, that is) to allow for an eventual redistribution.[45]

Here we focus strategic hiring on Categories X and Y, where the bulk of the effort is provided toward research and teaching. This is a policy decision that we believe is well-motivated by the fact that needs in Category Z can best be addressed by other means (see below). We think it makes sense, too, for the model to extend the policy assumption of adding equal amounts of faculty effort (FTE strength) in Categories X and Y. Other choices could be made, obviously.[46]

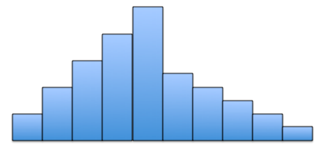
We estimate that to catalyze this transformation to a new steady state in ten years will require a surge of 20-25 new FTE (FTE strength). In modeling, it comes out of growth expectations for the data science education program, used as a general indicator of the expansion of campus-wide data science activity.[47] Over the scale of the campus, as we estimated to start, this corresponds to an overall sense of the research investment needed to strengthen capacity across this area.

If, for the sake of concreteness, we take 22.5 new FTE strength as our aim, the model with these policy preferences gives

Category X	100% effort	fully in data science	15 new headcount	15 new FTE
Category Y	50% effort	split attention	15 new headcount	7.5 new FTE

[44] Another simple policy option could be equal headcount in Categories X and Y. The reader may choose to develop a system of equations for the general case.
[45] Schematically, over the ten-year period, we need to build a new program that has some overlaps with existing programs. Our existing programs need to continue forward, and they also need to adapt. We can (must) tap into the existing programs, especially for teaching, but we cannot afford to undermine them overly along the way. At the beginning, our new added strength in data science will be small. The program's teaching, in particular, will come entirely out of existing programs' strength. As it grows, we need to bound that effect to let existing programs adapt. Then we will be through the surge, and we can phase the additional hiring back out.
[46] This choice to add symmetrically in FTE strength in Categories X and Y has some grounding in the education program model.
[47] The education program model assumes that as it trends out of a start-up phase relying solely on existing faculty, half the effort involved in delivering the varied pieces of the curriculum can be carried by existing faculty (shifting resources or multiple use of strength in existing programs) and half from growth.

For concreteness, finally, we can model this process happening temporally, say, with a five-year ramping up and a five-year ramping down corresponding to the nature of the surge process. Practically, this approach would respond to our growing capacity to field and staff strong searches. When it comes to transitioning back to regularly allocated positions, it has advantages over a plan that hits a cliff at the end. A purely illustrative hiring plan that distributed 22.5 new FTE (30 headcount) over 10 years might look something like this, for instance. Even in a peak year of, say, 5 new FTE, the surge would amount to a small fraction of total faculty hiring campus-wide.



This is a model. It is for illustrative purposes. It shows that a principled model can be built, and it gives sensible results. A great advantage of modeling is that identifies key assumptions and choices that can be discussed, and it prototypes a temporal plan against which progress can be tracked.

Character and scale

There are important questions to consider outside of formal modeling of capacity. With surge FTE available, we create the capacity to bring about targeted qualitative change:

- **We can choose to build a definite number of clusters of strength over time** as a means to achieve critical mass in these areas. For instance, we can deliberately target some subset of new faculty hires to seed a chosen area (for example, visualization or quantitative social science) by providing a center of attraction for graduate students, research funding, and potentially other faculty collaborators who can be brought on in later years.
- **We can choose to enable multiple faculty to be hired in an area in a year.** Their recruitment and addition to campus can put us on the map in an area and send a signal about the university's ambitions. Even the announcement of such a plan could be a magnet for clusters of faculty who might be attracted by the opportunity to be part of creating a new emphasis at Berkeley. If we choose to do cluster hires, our resources are circumscribed by the total number of positions we can make available, so such a move must be strategically planned. It might sensibly happen during a window of opportunity partway into the surge, at a time when there has been sufficient preparation, but before we have built enough strength that we could succeed without it.
- **The surge needs to be conceived in conjunction with the organizational structure** so that it simultaneously builds a strong core and strong connections. It needs to provide depth and breadth while creating synergies that enrich not just computing and data, but the entire campus.

Key observations:

- **Growth in the group of other, less involved faculty, Category Z, cannot be realistically catalyzed via targeted hires**, but would follow based on momentum from the investment and shifts in disciplines. A range of mechanisms are possible. If the campus chooses to communicate to departments that it will look more favorably at the margin on FTE proposals with a data science element, it provides subtle incentives without bringing them directly into a targeted hiring process. Individual programs or schools may find the campus backdrop useful in securing philanthropic support for data science elements directly in their home domains. Recruitments of new faculty in all fields with some data science expertise in their portfolio will be more successful against the background of our investment and a recognized pillar of academic strength. We also expect that the overall investment and new recruits can have a rapid effect in shifting the research and teaching focus of significant numbers of faculty already on campus.

- **Surge positions should not be thought of as *all* the hiring the campus will do** in fields in, drawing on, or related to data science. Rather, they make up the strategic part that is above and beyond regular allocations, some of which will likely end up in data science. Departments should not feel constrained to seek out the faculty they think are best for their programs, whether or not they are part of a surge.
- **There are no rigid assumptions built into the model about where these faculty are housed.** It would be sensible that Category X, fully in data science, would be housed inside a core academic unit (such as the School we recommend) responsible for computation and data, but that is not a constraint on the model. Category Y, split attention, could be inside the School, outside it, or joint.
- **These faculty positions can come from multiple sources**, including the regular campus pool, additional lines funded by philanthropy or revenue streams, or a mix of these. We expect that a mix will be the best option, acknowledging the constraints on the campus pool and the opportunities for outside lines, while making clear that data science is not some kind of foreign add-on to regular campus priorities.
- **There is no assumption that the total number of faculty positions on campus increases by the surge number forever.** That will depend on their funding mechanism (for instance, philanthropically endowed FTEs vs. other) and on policy choices.

The effect of this plan is to deploy a small investment to catalyze something bigger, a transition in the university at large. An investment of 20-25 FTE strength (and somewhat larger headcount) over ten years can feel like a large investment at Berkeley, where we have done relatively little concerted academic planning since the four New Initiative Centers were seeded with about this number of FTEs more than a decade ago. It is not large on any scale across other universities that are launching programs in data science. What will matter is that we do it truly strategically. Overall, as a surge investment, it amounts to 1-2% of our overall faculty strength. Yet over ten years, with good coupling mechanisms and enabling effects, it can be anticipated to bring us to a state with 250 faculty with some research and teaching focus in data science.

It would go beyond the FAB's mandate to devise a detailed plan. That will be a matter for strategic dialogue among the several levels of the campus administration, the Academic Senate, and the faculty. We would contribute several observations:

- Of the options available, a set of search models for Category X, fully in data science, could likely use both regular searches and joint appointments between key departments.
- A set of search models for Category Y, split attention, would take careful design (e.g., regular searches, half-FTE inducements, joint appointments, etc.) and would depend heavily on decisions about any kind of strategic focusing.
- We should assess our past experience in joint hiring, specifically identifying circumstances and practices under searches and appointments have succeeded (programmatically and for the individual) and those under which they have failed.
- Teaching faculty are a natural part of this model as departments or decanal units see a need.

Reflections

Though the FAB does not have access to FTE requests from departments, our experience leads us to conclude that programs in diverse parts of the university are already looking to bring in faculty with these mindsets, whether or not they always label them "data science." Commonly, Berkeley's search strategies for faculty have been shaped at the department level, rather than cross-departmentally by methodology. This process has been responsible for strong hires that are well-embedded in their departments. At the same time, it has created some unusual circumstances in which multiple departments have been interested in the same candidate. In recent cases, data science candidates have been highly ranked in multiple departments' simultaneous searches, creating situations that have been resolved either by informal arrangements or, in fact, competing offers. Data science is an opportunity for strategic academic planning in a well-defined and cooperative context.

Any strategic investment in faculty strength involves close collaboration between the Academic Senate and the administration. The design effort does not have to follow the FAB's particular model, but we urge that it be done on a principled basis. It will involve shaping a mix between lines from from the regular campus pool and new faculty FTE funded by philanthropy or revenue streams. Campus experience with FTE supported by the latter means has been limited, and in some cases it has been challenging. There is, however, we feel, a growing sense across the campus that Berkeley must learn to do this kind of planning, and do it well on a collective basis. The balance, timing (onset based on availability), and realistic incentive structures for units that try to secure philanthropically and revenue-stream-funded faculty positions need to be integrated into campus-level planning. This is a soluble problem that should get early attention so that it does not slow us down. The effort then should be continually monitored for lessons to be learned. In particular, campus-level allocation strategy for this shared resource must be designed with careful thought to mechanisms to mitigate common pathologies.[48] Here good practice will matter beyond data science. It may be be an important way to generate some level

[48] An informal list would include gaming by individual actors or abandonment of strategic decision-making by the center, such as degeneration into a lottery, a spread-the-peanut-butter-thin strategy, or capture by particular interests. We urge that the campus also learn from past procedural experience in strategic hiring, particularly so that we create processes that are not so unwieldy or painful to execute that participants are tempted to compromise on a sub-optimal outcome just to bring the exercise to an end.

of shared trust in our overall process as the university adjusts to the new financial environment in which it must operate in order to sustain its faculty excellence.

9. Strategic Foci in Domains

The preceding section of this report outlined a model for a campus strategy for faculty hiring that can support data science strength inside and outside the core unit. We see particular opportunities in exploring some new form of clusters of excellence that tap into the “broad” dimension of data science. Such strategic foci can bring together faculty who are already at Berkeley with potential new additions, who may find the university a more attractive home in the presence of a coherent data science effort. This coherent effort could include the existence of one or several marked-out foci for faculty FTE, as well as a strong core unit with connections campus-wide and a concentrated effort in fundraising, including for endowed faculty positions.

Academic units across campus are working with data in new ways, from public health and law to business and education to the humanities. On the FAB our sense is that three large domains offer particular opportunities at the moment; we explore each of them below. However, we by no means suggest that this list exhausts the possibilities. We strongly urge that a conversation around strategic foci for data science hiring in the “broad” dimension be initiated with deans, department chairs, and faculty. We urge, further, that this process be continued for the duration of the hiring surge, in dialogue with faculty leadership within the core unit, so that emergent areas can be identified and experimentation can be encouraged.

The three areas described below also give concrete illustrations of important synergies and benefits that are apparent to faculty from different parts of the campus if Berkeley proceeds to constitute a new school. Each of the opportunities for strategic foci that we examine below offers distinctive features around modes of institutionalization. This is natural and intrinsic to data science.

- Social sciences
- Biology
- Environment

Social sciences

Data science has a long history in the social sciences: a significant portion of the history of statistics is the history of the systematic collection and analysis of demographic and economic data. Every advance in data collection has transformed the social sciences and related professional fields, such as the post-war explosion of national accounts and randomized surveys. The current era is a pivotal moment because of the unprecedented explosion of detailed data on human beings and their environment. The granularity, scale, high-dimensionality, and heterogeneity of modern datasets are changing how social scientists do their work and what questions they are able to ask. Datasets in routine use in industry and government, which are increasingly available to researchers, contain information about a significant fraction of all human beings, enabling "personalized" inferences. The need to analyze data at unprecedented scales and velocities is drawing the social sciences closer to both the statistical and computational sciences. Researchers and students need to integrate a diverse set of skills, incorporating domain knowledge with inferential and computational thinking.

Data availability and computing have always been a critical constraint on the progress of the social sciences, with the availability of resources steering the direction of inquiry. For example, Gabriel Tarde, who conceived of social physics to be based on small interactions among individuals, "lost" to Durkheim because Tarde did not have access to the social network data needed to test his theories, and Durkheim's vision of sociological categories could be measured with data that was available at the time. The emerging availability of vast real-time social network data is making us rethink the role of such network forces and the definition of social facts. Troves of unstructured data, such as text, pictures, and video, represent massive repositories of data that is potentially transformative for the social sciences, but only beginning to be tapped. Integration of of diverse and high-dimensional data sources, such as overlapping economic, demographic, political, and environmental data, are generating new interdisciplinary fields that require innovative educational paradigms. The growing use of machine learning and field experiments are challenging traditional notions of causal inference, forcing social scientists to rethink established methodologies. Extraordinarily granular data sources, such as individual-level social-media behavior, and fully global data sources, such as real-time satellite imagery, open new doors to social scientists attempting to understand how macro-level phenomena emerge from micro-level motives. And the growing use of data-driven decision-making, both in private and public sectors, indicates the tremendous influence exerted by data science on modern societal outcomes.

It is difficult to predict where these new trends will lead, but we there is a clear need for some reorganization. Progress in individual social science fields has been accelerating, aided by the explosion of data and methods. But in many cases, data science work in a given social science domain field makes progress in relative isolation from the larger data science community, with practitioners in each field engaging primarily with researchers and students in the same field, relying on data sets and techniques that are widely understood and accepted within that community. Often culture, training, and methods allow these disciplinary boundaries to persist, even though data scientists in each field often have a great deal in common and might benefit tremendously from greater interaction with one another.

Berkeley is already home to leading researchers and programs in social science and data science more broadly; and Berkeley is an institution where cross-pollination of ideas and interdisciplinary exploration is standard practice. So it seems clear that Berkeley should lead the world in bridging divisions between data scientists in social sciences, professional fields, computers science, and statistics. Each group of researchers has ideas, methods, tools, and perspectives that it has developed and that would strengthen the larger community; and each group has room to learn and benefit tremendously from deeper engagement and understanding the other fields. At present, no university has "cracked the code" on how to establish and support fluid engagement between these various intellectual communities. In cases, individual researchers have found ways to navigate these divides, often having substantial impact, but this is generally in spite of, rather than in alignment with, existing institutions.

Nonetheless, given the history of data and methodological innovations disrupting social sciences, it appears likely that whatever community solves this institutional challenge of integration will be a world leader in the next generation of social science. Berkeley has extraordinary faculty leading cutting-edge research in data science in all of these fields. It also should be the place where cross-pollination between these communities, and structured interdisciplinary educational programs, accelerate all of them even further beyond the frontier.

Structured integration of the broader data science community with professional school programs and research is essential to their continued success. As the use of data science in public and private sectors becomes increasingly sophisticated and widespread, it is critical that our professional training programs and research reflect this advance so that our students and faculty continue to lead in these domains. Early investments along these lines have paid off. As one example, Berkeley's Goldman School of Public Policy has hired early and heavily among leaders in the use of data science for social science policy applications and is consistently ranked the top school for public policy analysis, the field in which data analytics are applied for the public interest. Deepening and broadening the connections between Berkeley's community of data scientists and professional schools is key to their continued global leadership and impact.

We recommend that the university undertake specific strategies to develop and strengthen the linkage between the social sciences/professional fields and the broader data science community. Our recommendations are framed as an invitation for Berkeley's faculty and administration to explore the options we call out, using mechanisms of consultation and a structured effort to convene a group of faculty who could choose to carry this process forward. Other options, of course, may get traction besides the ones we recommend; what matters is that the dialogue begin in the context of the larger effort that the FAB advises.

What would be needed for success?

Data science innovation is alive in Berkeley's core social science departments, including Economics and Political Science, but substantial data science innovation is also occurring outside of these core communities with no central "home" where multi-disciplinary innovators can learn from one another, share information, collaborate, build a community (including making hiring and tenure decisions), and produce a coherent voice for quantitative social data science at Berkeley. It is already the case that many data scientists in economics, political science, psychology, sociology, business, public policy, the School of Information, and other communities look more like one another than they like their disciplinary colleagues. Institutionalizing a "reactor"-like environment for these individuals to innovate and build a community would be a visionary and the first-of-its-kind step to establish Berkeley's leadership in the next decades.

Therefore, we advocate for the creation of a Program or Department in Interdisciplinary Quantitative Social Science, housed within the new school, if a body of the faculty in question agree. Highly creative individuals are crossing disciplinary boundaries of the social sciences and other data sciences, finding intellectual value in an integrated approach. The School should develop institutional support for those individuals already at Berkeley wishing to become more engaged with other fields and provide the institutional backing needed to attract emerging scholars in this new field. Such a program would be deeply connected with existing programs and would draw on faculty from around campus, serving as a conduit for collaborations and engagement across disciplines that house data scientists, while simultaneously providing an institutional foundation for the more unified community of data-driven quantitative social scientists that is emerging as a signature advance of the twenty-first century.

Success of this program would require:

Commitment of FTE and startup for future hires.

- Such a program should be "seeded" by leaders in interdisciplinary quantitative social science already present on Berkeley campus, but building a community of core faculty will also require dedicated hires from emerging and innovative programs around the globe.
- Joint or dedicated appointments for methodologists from statistics and computing, with strong social science interests, would provide a sound backbone for this technical community.

Educational programs.

- Educational linkages should be developed by supporting graduate and undergraduate students in social science and professional programs to navigate and access courses in other data science fields, and visa versa. Often programmatic requirements and faculty advisors steer students along well-trodden paths of methodological training, which is necessary for communication and credentialing within a field but does not facilitate cross-pollination across disciplines. Institutionalizing guidance and and programmatic linkages across departments can help students locate and obtain credit for appropriate training that will support their innovation within their home discipline.
- A dedicated graduate program can produce professionals, at the masters level, and intellectual leaders, at the doctoral level, who are comfortable and proficient at navigating data-driven research across all social science domains with a computing component substantially more developed than existing programs on campus. Flexible social scientists, fluent in multiple disciplines, statistics, and computer science are already highly demanded in both professional and academic settings. Training would be highly integrated with existing social science programs.

Fundraising.

- Much of modern quantitative social science has emerged from the data and methods developed in Silicon Valley, the location of many of the world’s most prominent consumers of this emerging fields’ new insights and innovations. Some of Berkeley’s

faculty in this area have strong connections into this industry ecosystem.
- As a key component of the new school, the Quantitative Social Science group will be a community that applies data science methods developed in the school to important social problems, making it a highly valued element in any competitive fundraising environment.

Biology

Biology is a field that has long been characterized by the systematic collection and classification of data. This aspect of the biological sciences has intensified in recent decades as a range of new technologies have emerged for collecting data, resulting in a series of “revolutions” defined in part by the kinds of data that have become available. Thus, not surprisingly, biology has been both a major driver and a key participant in the broad spectrum of activities that is now referred to as data science. This is true most notably in molecular biology, where technologies such as DNA sequencing and mass spectrometry have brought massive new data sets into the daily practice of biology, and in the exciting field of computational biology for protein structure prediction and drug-protein interactions. It is also true in neuroscience, where new technologies have been developed to collect brain-scale anatomical and functional data; in ecology, with advances in sensor technology and deployment yielding environmental data at a range of spatial and temporal scales; and in medicine, which is being revolutionized by a range of new technologies for measuring genotypes and phenotypes across populations, making possible innovative new approaches to treatment that are being referred to as personalized medicine.

Interestingly, these trends have not only increased the degree to which computational and statistical tools are incorporated into biology, but have also led to the emergence of an intellectual core, often referred to as computational biology. Drawing on ideas from computer science, statistics, molecular biology, mathematics, chemistry, bioengineering, electrical engineering, physics, evolutionary biology, etc., computational biology is making its own mark on the academic landscape, defining new areas of research that directly target emerging problems in biology. Rather than being merely interdisciplinary, computational biology has been referred to as antedisciplinary: it represents the first steps in the emergence of a new discipline.

This perspective was clearly understood at Berkeley in 2003, when computational biology was selected as one of the Chancellor's New Ideas Initiatives, leading to the creation of the Center of Computational Biology (CCB). Moreover, recognizing a unique Berkeley strength in the juxtaposition of sophisticated mathematical theory in evolutionary biology and population genetics with new capabilities for genome-scale measurement, a significant number of the hirings in CCB have focused on genetics and genomics. Core faculty participants in CCB have brought Berkeley to international prominence in computational biology. They and other members of the CCB have created a PhD program to meet ongoing strong demand and to begin to realize the antedisciplinary promise of computational biology.

This prescient initiative is foundering and is currently in a critical state. Several key CCB faculty have left Berkeley, and morale among faculty in the program is low. The reasons are clear: CCB has not been provided with space. Their original allocation of 7 FTE has not yet been fullfilled, and FTE requests for the past three years have not been met. CCB has not been placed within a college or school and is thus without a strong advocate at the decanal level. Faculty in CCB are partially in CCB and partially in other entities (departments), and in the context of current organizational arrangements, the dual sets of demands on these faculty have proved overwhelming. There has not been a way to explore whether the community might be placed on a path towards departmental status. While the number of applications to the computational biology PhD program is large, the admissions rate is quite small, and Berkeley does not tend to win in admissions battles against the universities that are strong in computational biology (Harvard, MIT, Stanford, University of Washington, etc.).

The declining health of computational biology at Berkeley contrasts with the striking growth of computational biology at peer universities, where our competitors (such as Stanford and Harvard) have launched new PhD programs and departments over the past few years. It contrasts with the high level of demand for computational biology classes by our undergraduate students and with the increasing need for computational biology expertise in our research labs, where there is an explosion of biologists performing high-throughput experiments who need help with modeling and data analysis.

We do not feel that failure of computational biology at Berkeley is a foregone conclusion, only that the window of opportunity for rescue is rapidly closing.[49] To the contrary, we believe that the school proposed in this document would create an intellectual home for CCB, allowing it to concentrate the people and the resources necessary to regain leadership. We focus on this argument in the remainder of this section.

As we have emphasized elsewhere in this document, the intellectual agenda for the proposed new school includes, but it is not delimited by, data science. Computational biology provides an illustrative example. The intersection of computer science and statistics that provides a basic point of departure for data science (narrowly defined) is highly relevant to computational biology; indeed, core computational biology involves complex algorithms that manipulate data structures (e.g., strings, trees, networks) for the purposes of making inferences (e.g., predicting future outcomes, inferring latent causes) in the setting of large amounts of data. And yet many key problems of computational biology involve concepts in the underlying sciences of biochemistry and biophysics, as well as allied fields such as epidemiology, immunology, and ecology, that have their own history and logic independent of data-analytic concerns. As time goes on, one can imagine computational biology as both remaining closely tied to core data-science areas in

[49] A further observation bears making: While the scope of computational biology is large, the scope of data science is far larger, with respect to a range of objective and subjective criteria, including demand from industry and the range of existing disciplines that are touched. Hoping that a modest investment of the kind made for CCB will suffice for data science is not realistic. The CCB has clearly fallen between the cracks at Berkeley, and it provides a counterexample for alternative proposed organizational structures for data science at Berkeley.

computer science and statistics while also developing its own unique agenda. Such co-evolution and unique evolution is of course true of existing schools and colleges, where constituent departments synergize but also develop individually. The point is that for the past decade, and for the foreseeable future, the alliance of computational biology with computer science and statistics is a very significant one. Placing computational biology in a school that contains computer science and statistics is clearly a win-win proposition, from many points of view, not only those of research, but also of fundraising and education.

Computational biology also illustrates the desirability and the achievability of a porous organization for the new school. Computational biology must continue to draw on domain expertise in the allied disciplines of biology, physics, chemistry, natural resources, and public health. Such connections already exist at Berkeley at the level of individual faculty, and by bringing CCB into the new school such connections are inherited. They can also be strengthened. Allowing the new school to be a division of the College of Letters & Science would allow the dean who is responsible for CCB to interact directly with deans from the physical and biological sciences, such that cross-divisional support can be provided for initiatives that tie computational biology to the sciences. Similarly, the proposed connection to the College of Engineering would allow cross-divisional support for initiatives that tie computational biology to bioengineering and other areas in engineering.

In summary, in addition to the arguments presented elsewhere in this report that justify a new school in terms of its effect on the computational, inferential and social sciences, we feel that an important additional argument can be found by considering the current state of computational biology at Berkeley and its critical need for an institutional home.

What would be needed for success?

Commitment of FTE and startup for future hires.

- Without computational biologists interested in human genetics, the new initiatives and efforts in precision medicine and the IGI will not realize their full potential.
- Currently we lack leaders in the field of computational structural biology, harming our prospects for synergy with our excellent structural biologists.
- With all of the excellent molecular physiology and "omics" at Berkeley, we need more strength in computational systems biology.
- Additional areas should be identified by broadly soliciting input from interested faculty, as the import of computation and data across many more fields of biology is growing.

Development of a functional core for computational biology.

- The current lack of funds for centralized computing structure and consulting staff has made it difficult for biologists on campus to benefit from the expertise of computational biologists on our campus. The proposed School could provide an intellectual home for a computational center as well as a full time statistical consultant employing students/part-time staff to assist biologists on projects.[50]

An undergraduate program.

- The number of biologists interested in computer science, statistics and mathematics (and conversely the number of students in those areas now interested in biology) means that there would be a significant clientele for such a major (or minor) and there are certainly an array of courses that could, and should, be developed and offered.

Fundraising.

- We have completely failed to take advantage of the increasingly important role of Silicon Valley in biotech, while UCSF, without our strength in computer science and statistics has done tremendously well at fundraising in this area.
- Under the umbrella of the new school, we will be in a particularly strong position to compete well with other institutions in local, national and international spaces.

Environment

Berkeley is a leader in research on the environment, with world-class programs from the physical, chemical, and biological environmental sciences to geography and agricultural policy, to environmental management and energy economics. These programs are generally dispersed, with only loose linkages to one another. Yet a unifying theme across these varied fields, many of which are highly interdisciplinary already, is their heavy reliance on data and their recent transformation driven by data availability, computing power, and methodological advances in data science.

Data on the environment has exploded over the last decade, as instrumentation of the natural world has dramatically expanded our ability to observe the world in which we live. This is true across many areas, whether through diffuse sensor technologies, satellite data, or field observations with data and images for millions of biological specimens, historical and recent, with associated ecological and genomic metadata. The ability to collect, aggregate, analyze and understand these massive data sets is now critical to all fields studying the environment and its management. For example, Berkeley climate scientists analyze millions of observations to measure and model the ways in which the global climate is changing. The Berkeley Initiative in Global Change Biology has developed a means to integrate vast amounts of disparate data to address the challenge of identifying the interactions and feedbacks between different species and their environment, and hence predict how biological communities will respond to change. Berkeley economists have analyzed of millions of household electricity billing records, helping the California government understand the effects global warming will have on peak energy demands due to heightened air conditioner use. Berkeley hydrologists have analyzed the implications of these warming trends for drought and water management throughout the West. A common denominator is the need to integrate data across multiple disciplines and analyze that data within a common framework.

[50] There is a highly successful system like this in place at Cornell University.

What would be needed for success?

Growth in environment-related data availability will continue to be rapid, and both employment for students and impactful research by faculty demands increasingly sophisticated data science tools. For these reasons, there is high demand for data science expertise, training, and innovation in environmental fields among students and faculty. In some parts of this diverse domain there can be a mismatch within programs between the training students now demand and the tools that faculty offer to teach, as well as uneven expertise among faculty across different subfields of environmental study. Thus we see extremely high potential for an expansion and integration of interdisciplinary data science training and research across environmental fields.

Data science is emerging as a methodological connective tissue for the Berkeley environment community which shares a common interest but which is otherwise highly diverse and diffuse. Integrating and connecting environmental programs with one another and with core data science disciplines will both empower individual programs, by exchanging critical training and expertise, strengthen the broader interdisciplinary environmental community across campus, by creating a common language and supporting regular cross-disciplinary interaction, and amplify the impact of all Berkeley's environmental work, by grounding it in cutting-edge data science and providing common institutional support. We have local examples already nucleating. For instance, beginning its second year, Berkeley’s innovative NSF-funded National Research Training program “Environment and Society: Data Sciences for the 21st Century” (DS421) represents a testing ground for these ideas and is demonstrating the effectiveness of using data science graduate training to connect individuals and disciplines within the environmental domain.

The intersection of environment and data science on this campus will continue to grow and change. Development of the core school envisioned here offers the possibility of catalyzing more rapid adoption of state-of-the-art methods, developing a unique data science of the environment, and strengthening Berkeley’s preeminent position as a leader in environmental science and policy. In contrast to the two other domains we review in this section, we believe that the intersection of data science and the environment will benefit from a more distributed arrangement of faculty strength, an ongoing process of bottom-up choices in relation to faculty hiring, and a locally driven process of fundraising that can leverage and align with the overall campus pillar of broad data science strength.

10. Fundraising

Active and successful fundraising is both a requirement for our data science enterprise and a key opportunity for the campus more generally. Computation and data analytics offer significant opportunities to develop Berkeley's capacity to draw upon philanthropic resources. Steps taken by the university to draw together computationally centered or data-intensive fields will increase their philanthropic visibility; connecting them with application domains will add to the intrinsic appeal on both sides. The FAB sees data science as an outstanding example of a widely diverse set of fundraising opportunities that can elevate levels of support in different parts of the campus. We believe, too, that a commitment to the Berkeley vision for data science, with its deep, broad, and rich aspects, will help us shape a fundraising strategy that remains in line with our university's fundamental public mission, at a time when external resources will be increasingly important for that mission to thrive.

Recommendation

Move data science rapidly forward as a central pillar of fundraising across Berkeley, including new endowed faculty FTEs inside and outside the core unit, support for key programs and institutes, and the new School as a whole. These elements should be coordinated in an overall plan; securing them need not happen all at once for the overall process to succeed.

Experience at Berkeley and elsewhere strongly suggests that this is a powerful philanthropic avenue to pursue. The robust participation of Berkeley's computing and data science faculty in the industry ecosystem gives us much leverage, creating opportunities to invigorate fundraising practice university-wide. As part of the campus's new strategies around fundraising, this pillar can be constructed in a coordinated fashion that bridges across academic units. We see major benefits for other academic units seeking to raise funds in data science-enabled application areas to be able to align with this university-wide pillar.

It is possible to sketch philanthopic opportunities only in broadest outline in this report. What matters more is the overall approach. At the macroscopic level, given the broad success of other universities in developing well-grounded philanthropic models for large-scale data science initiatives, it would seem that the signals from the national landscape can only be called positive. This macroscopic picture will need to be understood in the context of Berkeley's local strengths and weaknesses in fundraising for data science-related fields and ventures.

Berkeley has been technologically and geographically at the heart of the largest, most rapid accumulation of wealth in human history. This context is only visible to a very limited extent in our fundraising portfolio. There are several strategies that the university would need to implement, together, to make this fundraising effort successful:

- A pivot toward developing philanthropic relationships with the computing and data-analytic communities. These individuals are somewhat different from many of the donor communities who have given generously to Berkeley in the past. They will require both additional outreach resources and modifications to our usual approaches.
- A strong, Chancellor-level commitment toward large scale fundraising for this purpose. A $100M-scale gift would change how this effort develops across campus. This includes making data science a visible priority in the forthcoming campaign.
- In some past cases, e.g. the Hewlett chairs, a large fund has become available for allocation across the campus. A similar program would be a game changer for data science because it provides a natural source of coherence.
- It is also necessary to pursue philanthropic contributions at the scale of individual endowed faculty FTE, using models that will need to be developed and applied across different domains. These will need to be coordinated with the opportunities for strategic foci for hiring in targeted domains, as sketched above.
- Fundraising for faculty support in data science will likely require greater coordination, so that a faculty member with a split appointment can get coordinated support (central and decanal) for fundraising around the faculty member's research.
- Gifts in support of programmatic elements such as the data science curriculum, BIDS or other institutes, intensified measures for connectivity, or focused efforts to address questions of data and the public good will each require a distinct strategy.
- The campus should recognize that it will need to embark on fundraising for the new school as a whole, with the possibility of a significant naming opportunity, and for the possible creation of a physical home, including a building. These possibilities will need to be developed, of course, in closest consultation with university development.
- Manifestly, all these efforts must be done thoughtfully and strategically, in line with good campus practice around costing and long-range sustainability.

A collective understanding of fundraising for data science will also, we stress, make it possible for the many units across the university that are seeing upsurges of data science activity on their home terrain to develop philanthropic messages in relation to a broad campus-level emphasis. Fundraising that is done in relation to a campus-level pillar will intrinsically benefit each distinct unit that is enabled to secure a gift. In the campus's exploratory experience so far, it seems very likely that the "broad" and "rich" dimensions of data science will appeal to diverse philanthropic audiences who may not have interest in giving to "deep" domains. This approach maps on to the spread of data-analytic approaches across many sectors of business, opening new opportunities for endowed chairs across the campus: from friends of the university who want to give back to their domain of training, but acquired their wealth and giving capacity via harnessing the power of data science in their respective businesses.

There is reason to expect that a new school for computation and data at Berkeley will be a powerful attractor for philanthropy. However, there is no sense in which a core unit will “own” this area. Distributed fundraising across the campus will take cooperation and coordination among central, decanal, and other operations. Where this has sometimes been challenging at Berkeley, it points to a fundamental need in our shared future for a more collaborative approach, and the leadership of the new school must be prepared for and committed to this strategy.

To gain an overall and a granular sense of the possibilities, a necessary step for proceeding is to create a data science initiative whose potential elements can be discussed with advisors, industrial partners, and possible donors.

11. Revenue Generation

Given our diversified strength, there are many opportunities for revenue generation in data science at Berkeley. We expect that deans, including the dean of new school recommended in this report, will hold responsibility for planning, securing approval for, and fielding whatever programs they seek to offer. The FAB aims only to indicate some of the possibilities apparent to us in this space, with a partial but not exclusive focus on those that align with the assets and programs of a new school. It must be underlined that revenue generation for data science will surely be as diverse and distributed an ecosystem at Berkeley as data science itself.

Revenue generation in educational offerings

The transformations we have discussed in this report with respect to the university are even further along in the larger industrial ecosystem. The need is not just for educating new entrants into the data-intensive workplace, but recognizing the needs of generations ahead of them to re-educate themselves to attend to these changes. This suggests a spectrum of revenue generating opportunities. It is in every way in Berkeley's interest to move emphatically into this area, given our reputation, our diversified strength, and the platform that our growing educational efforts can provide.

Expanding professional masters capacity

There is substantial growth potential in the intermediate tier of professional masters programs. As outlined earlier in the report, Berkeley today has several masters programs that prepare students for different aspects of the professional practice of data science. Some of them would very likely transition into the new school; others may very well remain outside. There are multiple other masters programs that we understand are in the stage of being envisioned or developed by academic units across campus with strengths in particular fields. The FAB has not seen it as its job to evaluate existing masters programs. We do observe that there is no reason to expect there to be a single, unified such program. Rather, there is value in having multiple flavors of such programs, reflecting the different characteristics of the many distinct job markets in which data science plays a significant role.

Even with our existing offerings, there are gaps that are likely to be filled, either with new programs or expansions of existing ones. Saliently, no existing program focuses on training professionals at the core of the field, building out data science infrastructures. Computer Science and Statistics may want to team up to offer a program focused deeply on systems and statistical machine learning. Further afield from the core unit, many institutions offer a business analytics focused masters, as Haas may wish to introduce in the future. It is possible to envision many more.

Professional development besides the masters

While we typically view the professional masters as targeting those interested in expanding their technical skills in order to make a qualitative shift in their career track, there are equally important professional needs for which the masters may not be the ideal tool. One example is professionals who are on or shifting to a range of distinct management trajectories who will need to be able to lead effectively a team of professionals in varied sorts of data-intensive workplace and make good recommendations to upper management. Addressing these needs within the envelope available for such professional advance may mean creating certification programs that are more compact than a masters, more oriented on decision-making, and less on gaining technical proficiency. Even within the technical ranks, specific targeted coursework on a scale much smaller than a masters serves an important function; and various on-line options make it more viable within the work schedule. Similarly, executive level professional are increasingly faced with making good business decisions in an increasingly data-centered ecosystem. This speaks to a range of executive education opportunities that have a pace, duration, and character quite different from the avenues above.

Building stronger educational connections to our industry ecosystems

Given the recent emergence and broad character of data science, the opportunities for career placement for our students in this domain are quite different from the traditional close association (for instance, in engineering fields) of a department and major firms and industries it serves. Many companies and many non-IT industry sectors are hiring data scientists of various forms. They face a profound challenge in trying to reach appropriate segments of the undergraduate student body for certain kinds of entry-level positions, and also in cutting through the background of fast growth IT players. With the kind of institutional structures discussed throughout the report, we could envision the new school providing a natural point of contact for a broad industry segment and its connections as a way of tying industry needs to segments of the student body who would be interested and able both in data science and in the industry segment it serves.

This approach holds the promise of building more integrated career placement opportunities than simply supporting the recruiting process. Companies would have an interest in supporting the program on a regular, rather than episodic, basis to gain visibility amongst the student body. There are avenues to work with industries in preparing students to best represent themselves in the interview processes of various industries. Moreover, possibilities exist to build mentor networks into these industries that both draw students and support them on the job. There are co-op opportunities where projects within the industry setting contribute to expanding students' educational experience while gaining greater understanding of particular career choices. Capstone course opportunities are significant. All these options build connections that open both programmatic revenue sources and increased philanthropy.

In the burgeoning sector of immersive coding bootcamps and accelerators, outside the university setting, we have seen new models for interweaving education and professional development, including notions of taking equity in student growth through deferred tuition and other innovations. Data science may be a place where such innovations are considered.

Entrepreneurship

Many of the most important data science companies are yet to be founded; many will be founded by our students and faculty. While we have today at Berkeley several startup incubators and modest connections to venture funds, it is likely that the conjuncture of our growing data science programs with the formation of new industry opportunities will bring entrepreneurial opportunities very much to the fore. We need to approach this very wisely, paying close attention to issues of academic freedom, scientific integrity, and fairness. At the same time, we would hope to enable the university to help secure its future in connection with its role in creating entirely new industries.

Part III

12. Situational Challenges

We will need to come to terms with situational challenges (at Berkeley, at present) in order to realize the recommendations in this report. We need to acknowledge this reality and address how we can work with it. The FAB strongly believes that a realistic assessment of our situational challenges still points us to proceed.

We break down these challenges into five headings:

1. Resources for the plan and the transition, given Berkeley's current financial situation.
2. Campus-level strategic planning for faculty FTE.
3. Campus-level strategic planning for reorganization in general.
4. Absence of previous experience with a cross-college division.
5. Interdisciplinary inclusivity, flexibility, and adaptivity.

Resources for the plan and the transition

In order for the overall proposal and for the new school to succeed, it will be critical that both are supported with the resources needed to secure the transition and to foster growth. Here we use resources in a broad sense, including faculty, staff, and space, as well as research- and teaching-related resources. Some of the most important needs overall include resources to support the shift and growth in faculty size in the new school and across the campus, on the order of 20-25 FTE strength. Sufficient staff resources to support the scale of the school will be needed. Space to house the core groups and functions of the new school will also be critical.

We recognize that this process of transition and growth will happen against a campus backdrop of severe financial constraint. It is important that in the process of building and growing a new entity, existing programs are not handicapped and planned resources not subtracted from them. Through the transition we will likely be placing additional stresses on these units, and as a campus we cannot afford significant losses in their capacity. Some substantial part of the resources needed for the new school already exist within the footprint of the programs that are likely to move in. Another substantial part will need to be generated from outside sources.

We must start down this path by assessing and realigning the resources already at hand. These begin with current staff, space, and budget allocations for units that may join the school, as well as reserves and other resources that support faculty research and teaching. The principle that should guide this process is that resources should follow the faculty members and programs which they support. The work of restructuring may bring about financial savings through redistribution of resources and elimination of duplication, or it may not. It is premature to predict now. In either case, the amount is not large and is not a reason not to proceed. Success in this part of the process depends heavily on constructive collaboration with the decanal units where programs are currently housed.

At the same time, it is critical for all participants to understand that some of the resources to support the larger ambitions of the plan will need to come from new sources. As we have outlined above, we see considerable capacity for revenue-generating degree programs or other educational programs. Some of the units that will likely form the new school already have such revenue-generating programs in place. Others, including executive education and co-op programs, could be natural outgrowths of the new unit. Just as significantly, philanthropy will manifestly be a major component of any growth plan for data science, as we have sketched in an earlier section. Berkeley's need to understand more granularly *how* to secure philanthropic support in this area is a key reason to start exploring it immediately, no less important than actually securing support. The growth plan in data science will necessarily be paced in part by success in raising new resources. Here, too, success depends heavily on constructive collaboration with UDAR, with decanal units that are active in this space, and with the Vice Chancellor for Research.

Moving forward with this plan in the midst of Berkeley's financial challenges will take stringent attention to costs and considerable inventiveness. To enable the transitional process will require drawing on some campus investment, principally to cover staff time, while recognizing that this central allocation cannot be large and may in fact require raising philanthropic funds to cover part of the cost. We do need to invest immediately in new fundraising capacity responsible to the leadership of the effort and dedicated to the overall plan and the projected new school, including the transition period as well as the growth. Given the large potential both for success for the new school and for benefits for the rest of the university, the FAB concludes that this is a powerful investment that Berkeley must make even at a time of campus-wide financial challenge.[51]

Campus-level strategic planning for faculty FTE

In its practice of relying on regular mechanisms for faculty FTE allocation, Berkeley has not carried out a campus-wide effort of strategic academic planning for more than a decade. Some of the limitations of the previous round (which led to the four New Initiative Centers established in the early 2000s) suggests taking a different approach for data science. As a faculty body we have little recent experience conducting strategic campus-wide discussions about hiring. To succeed, our discussions for data science will require thoughtful dovetailing of multiple institutional actors and decisions based on general principles that stand up to scrutiny.

To align with shared governance, it will be critical to have strong leadership of this effort in both the campus administration and the Senate. The FAB wants to express confidence that Berkeley is in a position to take it on. We hope to have advanced the process by drafting a model (in an

[51] High-level considerations about transition planning are provided in an appendix.

earlier section) that allows us to highlight possible strategies for principled decision-making, a model that no parties are obliged to accept. We also express hope that the breadth and inclusiveness of data science, including the different ways that diverse units can leverage it to request or raise resources, will put the university in a good position to work out a campus-wide model that respects particularity even as it crystallizes out principles.

Campus-level strategic planning for reorganization

Procedurally, the pathway to organizational reconfiguration is clear. Today its outline follows the division of roles between the faculty, the organs of the Academic Senate, and the administration in the Berkeley Review Process Guide and the University of California Compendium.[52] However, the campus has not exercised this option at decanal scale since the 1980s. The reorganization of biology during that decade, in the limits within which it unfolded, was itself the largest institutional effort at structural renewal that our campus has made since the late post–World War II period.[53] The FAB shares the sense that the present is a moment like Berkeley's reorganization of biology in the 1980s, of similar significance to the campus. Yet as much as we see it as essential to do justice to emerging developments around data science and computation, we recognize that this form of change feels momentous at Berkeley, perhaps more momentous than elsewhere. For that reason, we underline that the challenge of reorganization is more than a matter of following procedure. It is about carrying out the process in a way that strengthens the outcome for the campus, meaning that openness and fair-mindedness are key.

The distinctive challenge for computation and data comes down to the forms of coordination and cooperation this configuration of fields requires: coordination among core and connection domains on the one hand, cooperation among deans on the other. It is true that Berkeley has not gone so far down this road of collective action before. Structurally, we are de facto a decentralized organization. The FAB believes that the organizational novelty of data science points to an opportunity for constructive change in our campus-level thinking. We see this as a key contribution to addressing Berkeley's challenges looking ahead, working out cooperative ways to do strategic planning for fundraising, for instance, and for the university's future. The FAB believes that the organizational reconfiguration we recommend can succeed when it respects organic processes of faculty decision-making and the procedural guidelines we now have and when, in addition, it is grounded on empirical data, clear definitions and principles, and

---

[52] Berkeley Campus Review Process Guide for Academic Programs and Units (2011), http://opa.berkeley.edu/sites/default/files/ReviewProcessGuide.pdf; Compendium: Universitywide Review Processes for Academic Programs, Academic Units, & Research Units (2014), http://www.ucop.edu/institutional-research-academic-planning/_files/compendium_sept2014.pdf.

[53] Data gathered by former Vice Provost Broughton from the online general catalogue indicates that since 1961, the same five colleges have been in existence at Berkeley, though sometimes with slightly varying names, and for most of the same period, the total number of schools has been nine. For the reorganization of biology at Berkeley there is no definitive historical account. We have gained insight from the perspectives captured in Martin Trow, "Biology at Berkeley: A Case Study of Reorganization and its Costs and Benefits," Center for Studies in Higher Education Research and Occasional Paper Series 1.99; Roderic B. Park, "Lesson 2 - Reorganization of Biology," in *It's Only the Janitor: A Handbook for New Academic Administrators* (Geyserville, CA: Rockpile Press, 2010), 119-144; and the recollections of colleagues.

well-defined memoranda of understanding that secure our relationships of collaboration and cooperation.

Absence of previous experience with a cross-college division

If Berkeley takes the less traditional route of forming a new school as a division in two colleges, we pay some cost in overcoming uncertainty and establishing new practices. While the FAB overall sees attractions to this model alongside the other possible implementation in a free-standing school, we think it is important to identify concerns that may derive from the absence of a previous instantiation. A bi-college division is an experiment at Berkeley. While we believe the model is structurally well-formed in its fundamentals, experiments often bring surprises in practice, which can feel unwelcome when it would feel more comfortable to be on a pre-defined path.

For instance, there are overheads in formalizing new college-level governance structures that would not be present with a free-standing school. Even if those costs are judged to be worth paying, they exist and introduce an element of contingency upon the reactions of the leadership of the colleges. Those reactions to the options proposed in this report are yet to be determined and thus a factor of risk. Although the main lines of resource allocation (faculty FTE, annual budget, TAS, faculty salary savings, and overhead return) come directly from the provost to each dean, and other key functions (strategic planning, academic personnel, and budget requests) go back up the same chain, there are shared resources in each college in the form of deans' offices whose use requires coordination. The likely overlap in fundraising also points to the need for explicit coordination around prospects and messaging. That said, formal coordination between deans will almost certainly be required for any significant fundraising to support data science, whatever the organizational form. Thus there are

There is also a level of administrative complexity around matrixed infrastructural models that cross colleges, in which MOUs are necessary to keep lines of resource provision and decision authority aligned. While we have models of programs that are supported in a coordinated fashion across colleges (for instance, the two major programs of study that are presently fielded by EECS faculty in the College of Engineering and the College of Letters and Science), we have rather less experience with units or programs that are located in two colleges simultaneously.

Possibly a more challenging element is cultural novelty, i.e., faculty, staff, and others feeling puzzlement at the implications of the new structure. While we believe that many of these concerns can be addressed by thinking about decanal units as primary and colleges as governance, admissions, and administrative structures constructed around them, we recognize there is a cultural challenge deriving from the absence of previous experience and a sense of identification with existing boundaries. It is our hope, and, for some of us, our expectation, that these concerns would fade as the operation of the unit became familiar.

Administratively, creating the cross-college division is certain ways more complicated than creating a free-standing school. It would require changes to the campus organizational chart and to routine practices around attributing the resources and activities of a decanal unit to one college or school. On the other hand, it is an easier transition if it is possible to continue to use existing programs' infrastructures in their current situations, rather than sever, move, and reconstitute them in a new context.

Conversely, it may be easier for faculty or leadership in the affected colleges to contemplate the transition if it is not tantamount to an exit. We expect some faculty will prefer to think of the cross-college division as an incubation model for a free-standing school that may come at some point in the future, depending on institutional evolution, cultural identity, and other factors.

Interdisciplinary inclusivity, flexibility, and adaptivity

Berkeley's experience is that interdisciplinarity is both exciting and challenging. The two sides of the coin are probably in the nature of interdisciplinary work, as an extensive literature on the subject suggests. Hard experience can lead to skepticism about the feasibility of interdisciplinary constructions, especially when these are promulgated without realistic attention to incentive alignment, resource constraints, or potential conflict. Doing well in this area will require careful assessment of experience local to Berkeley, where we have successes to draw from and experiments to observe, as well as distinctive opportunities and particular mechanisms as outlined in the section on connectivity above.

The FAB's sense is that in a future in which data science will be used broadly, it will indeed be necessary to figure out how to align incentives for faculty in these areas to succeed in their careers. We underline that the new unit must be broadly inclusive in order to achieve the synergies and cross-cutting interactions that are some of the greatest potential benefits of this restructuring. This poses a number of challenges, including how to draw certain boundaries so as to maintain intellectual cohesiveness, and how to maintain close ties with nearby units. A model where individual faculty can opt to participate in the intellectual agenda either as part of the core or of the domains of connection would encourage and strengthen inclusivity; we have addressed these points in the section on connectivity, as noted. We set this ground-up approach in contrast to top-down mechanisms to create interdisciplinary institutes that, despite our best efforts, do not generate the kind of synergies we seek here.

It is also challenging at Berkeley to design for flexibility and adaptability at the institutional level. While we have considerable experience in staying agile at the sub-department level (for instance, fluidity across divisions within MCB, forming and re-forming group structures in EECS), these have depended on fitting into an unstructured space in the academic organizational hierarchy. Data science changes on a fast time scale, however, and it is really a promissory note on a future that has not yet come into being. We are obliged to design for change: to do our best to create structures and practices that can support more than one possible configuration, to build in reporting and assessment practices that enable introspection, evaluation, and revision, and to allow for approaches that permit us to revisit decisions at appropriate intervals, such sunset clauses. These steps make it possible for us to be responsive to the ongoing evolution of data science, without assuming that it will go down a predefined path. The organization and governance structure for the new school will thus need to balance sufficient clarity and definition in order for the unit to get off the ground with clear expectations about many operational principles, with sufficient flexibility to allow it to adapt as it grows. The need for flexibility arises both because unit will need to be organic, shaped by those who join, and because as the field advances it will need to be poised to adjust to respond to new opportunities and changing landscapes in order to remain at the forefront.

13. Recommendations for the Process to Follow

The FAB concludes there is a clear path for Berkeley to follow to cross-cutting success in data science in a way that engages effectively with our already established programs. Ground-up efforts have brought us this far; it is now time to take campus-level action. We need to move rapidly to tap into and sustain faculty engagement. The path ahead is not short, and the destination is not predetermined, but the steps to embark on are well-defined. To recap, we urge the university to undertake a rapid set of interdependent measures.

1. **Organizational form: Move to create a flexible, innovative academic core of independent decanal stature**, a School centered on computation and data science, **with a mandate to develop a robust culture of engagement and strong mechanisms of connection campus-wide**. After considering multiple options, we recommend that the goal be to form this School administratively as a Division of two existing colleges (Engineering and Letters & Science), although it could also exist as a free-standing academic unit. Departments, programs, and institutes can use regular faculty governance processes to populate it; our strong sense is that a world-leading school can be built.
2. **Strategic academic plan for faculty FTE: Invest in an influx of data science faculty positions** (over 10 years, on the order of 20-25 FTE strength). Faculty positions are key to both expanding core domains and building broad strength as this area surges. **Provide a path for faculty to identify targeted application areas for decisive investment of FTE**. Immediate opportunities are in the social sciences in their intersection with computation and data science. We see the need for a next-generation strategy for data-intensive biology and significant possibilities around data science and environment; there will be other emerging areas as well.
3. **Fundraising: Move data science rapidly forward as a central pillar of fundraising** across Berkeley, including new endowed faculty FTEs, support for key programs and institutes, and the new School as a whole.

For this effort to succeed, all three measures are necessary and need to be taken together. To maintain the momentum we now have, they should begin without delay. Along the path of each of these actions there are decision points, with connections and dependencies among them. Multiple groups of participants are involved—the faculty, the administration (central and decanal), and the Academic Senate. That does not make the effort infeasible; it just makes it important that we manage it as a coherent process of intentional change.

Data science initiative

In advance of creating the academic core unit and executing a search for a dean, we urge the formation of a Data Science Initiative (or some such name) as **a transitional vehicle to operate for roughly 24 months**. This next-step initiative offers a way to mobilize faculty energy, provide executive function, and offer strong ground-up partnership to campus leadership, continuing the work of the Data Science Planning Initiative in engaging faculty, the administration, and the Senate.

We recommend placing the initiative under the leadership of a **faculty director** charged and empowered to move forward with these measures **in collaboration with a broader faculty team**. In this leadership structure we think it will be valuable to underline forms of collaboration and assistance by creating an academic committee of the heads of academic units that may become part of the new school, as well as an executive committee of associate dean-level representatives of academic administration from the colleges, the body of professional schools, and the Senate. We recommend explicitly authorizing the director to create additional faculty roles within the initiative and to constitute supporting committees such as a faculty advisory board.

We recommend defining **five goals for the new initiative**—three strategic aims, to be executed in high-level partnership with the administration and the Senate, and two interim functions, carried out with ground-level collaboration while a new organizational form is put in place. All of these are opportunities to broaden the engagement of faculty and units with the campus's efforts in data science. We lay out below one way of doing this, though there are of course other possible constructions. We recommend a fast-moving timeline that still gives sufficient time to do the work well.

Goals 1-3: Strategic aims of the initiative

Aim	Partnership model	Proposed timeline
1. Spearhead the process of **constituting a new school.**	Lead: Initiative director in partnership with academic and executive committees, in stewardship for the broad group of faculty involved. To be reviewed by Senate and administration (campus and system) as per BRPG and Compendium.	Complete constitutional process by June 30, 2018. Implement in AY 2018-19.
2. Define campus-wide **strategic academic plan** for faculty positions on a scale of 20-25 FTE strength.	Joint lead: Administration (VPF), Senate. Strategic input: initiative director, deans, and department chairs.	Complete plan definition by June 30, 2017. Implement starting AY 2017-18 if possible. Duration ~10

		years.
3. Launch and execute **fundraising efforts** at the necessary scale.	Joint lead: Initiative director in partnership with executive and academic committees, administration (UDAR). Strategic input: deans and Senate.	Implement at initial levels continuing directly from DSPI. Complete overall plan, with hinge points, by April 1, 2017.

Goals 4-5: Interim functions of the initiative

Function	Partnership model	Proposed timeline
4. Grow the undergraduate **Data Science Education Program** in both core and connection domains.	Lead: Initiative director with supporting faculty roles. Administrative support and oversight by L&S Division of Undergraduate Studies. Coordination with Budget Office on TAS model. Approval of Senate for creation of course and programs of study.	Continue expansion of program in depth and in reach. Strengthen the platform, connect to additional departments. Prepare for transition to new organizational form by June 30, 2018.
5. Strengthen and integrate **data science in research and graduate education**, in collaboration with other established programs.	Joint lead: Initiative director with supporting faculty roles, BIDS. Strategic input and support: administration (VCR, Graduate Division) and Senate.	Extend outreach in research. Generate network effects. Develop shared graduate educational strategy. Prepare for transition to new organizational form by June 30, 2018.

The initiative director will need to operate at the level of an interim dean of a new unit. **The role of the director is independent of and entirely decoupled from the role of a future dean.** The position of a future dean would be filled by the standard procedures of a decanal search that can be authorized when it seems likely that we are moving ahead.

Campus leadership next steps

Achieving the initiative's goals requires starting now. In particular, given the processes involved in academic organizational change, to get to the endpoint of a school even two years away, in 2018, it is essential to take action in fall 2016. Faculty commitment depends on **clear signaling of high-level administrative intent**—which requires not a statement that the university already knows how the whole process will turn out (which would get ahead of necessary dialogue and review), but a statement that the challenge is worth tackling, the goals are definite and are shared, and the transitional vehicle of an initiative has the administration's full support.

We thus recommend that on receiving this report, campus leadership (administration and Senate) and key stakeholders undertake necessary consultations in summer 2016. For the process to unfold further, the outcome of consultation needs to be a clear decision by the administration about the desirability of embarking on this path and the identification of the organizational option it can support for forming the school. It is necessary for the highest levels of the campus administration to communicate the decision with a statement of intent, articulating the goals and including rough parameters setting the intended scale for Goals 1-3. This will make clear the direction the administration seeks to go to the administrative and academic organizations it controls or oversees, to the Academic Senate as its partner in shared governance, and to the faculty who can shape their own next steps in its context. If campus leadership agrees, the FAB thinks it would be appropriate to share broadly the body of this report.

Constitution of a new organizational form (school)

The task that starts from the faculty is associated with Goal 1 (organizational form). Section 6 of this report outlines principles and processes for forming a new school. This is a **constitutional process**, and stewarding it effectively takes both **articulation of consensus** and **executive function**. The formal steps toward review and approval of a proposal for a school are documented in the Berkeley Review Process Guide and UC-wide Compendium.[54] Before these steps can be taken, there is considerable work to be done. This work can be shepherded by the faculty director of the initiative with the support of the academic and executive committees. They will be tasked to articulate directions, build consensus and momentum, coordinate faculty efforts to prepare aligned documents, and move them through the steps of Senate and administration examination and (we may hope) approval.

As in section 6, we recommend structuring the process by sharing information broadly, then beginning dialogue and consensus-building with faculty in departments and programs that can use their existing internal governance processes. Other emergent clusters of faculty may also participate if they can coalesce and give themselves governance principles on the same timeline.

This phase should not be rushed, but it should also not be drawn out. While it is ongoing it represents a period of uncertainty, which is distracting. Sufficient information and guidance should available up front so that the process may be swift; there should be sufficient time for deliberation, but also a clear schedule. It is particularly important that this not be an iterative

---

[54] Berkeley Campus Review Process Guide for Academic Programs and Units (2011), http://opa.berkeley.edu/sites/default/files/ReviewProcessGuide.pdf; Compendium: Universitywide Review Processes for Academic Programs, Academic Units, & Research Units (2014), http://www.ucop.edu/institutional-research-academic-planning/_files/compendium_sept2014.pdf.

negotiation process subject to gaming.[55] Commitment by a unit will need to make clear which faculty do not intend to participate with the unit so that alternatives can be formulated. Similarly, if a unit chooses not to participate it should identify individual faculty members who intend to do so. Given the declarations of participation, it can be determined if there is critical mass to justify moving forward with a proposal for institutional change.

Assuming appropriate participation, the next stage is the **concrete proposal process** documented in the BRPG. We recommend that this work proceed via forming a collective task force stewarded by the initiative that should produce a pre-proposal as per the BRPG and work to gain consensus among participating faculty. According to the BRPG, this pre-proposal would be submitted for review by the campus administration and then to the Berkeley Division of the Senate. Through these stages the deans have the responsibility to work with the faculty and the campus administration to produce a viable proposal and to work out relevant operational plans.[56]

Later steps at the UC level will follow as detailed in the BRPG and the Compendium. The school can first said to be ready to be constituted when the pre-proposal has been developed into a proposal that in turn has received all necessary approvals. In that sense it is only at the end of the road that Berkeley can make a definitive statement. However, **a well-framed process signals intention and makes clear what the decision points along the way are, allowing them to be coordinated with decision points for the other initiative goals**.

Within this constitutional process are significant matters of detail. What matters is that the spirit and terms of the process are clear. We suggest the following timeline.

Aug 2016	Statement of intent from campus administration (5 initiative goals), communicated to faculty, departments and programs. Stewardship of constitutional process begins by initiative leadership. Preparatory work of data collection and assessment begins.
December 2016	Statement of intent to participate by existing units or newly constituted groups, achieved through engagement of recognized governance practices. Formation of task force to draft pre-proposal documents. Pre-coordination with Senate and campus administrative offices.
February 2017	Completion of pre-proposal documents. Socialization within participating units and groups, and more broadly with campus. Revisions as required to lead to a shared set of documents approved by participating units and groups.
March 2017	Submission of pre-proposal for campus administration review. If feasible, forwarding

[55] In game theoretic terms, we seek an incentive compatible mechanism wherein it is in the best interests of the faculty to simply convey their true utility function.

[56] Much of the work required to complete a proposal will require assistance and support of the deans' offices, so this should be viewed as a cooperative effort. The BRPG specifies the approval authority; it does not suggest best practice for constructing a sound proposal and bringing stakeholders together. The latter is is essential.

	by administration to Berkeley Division of the Academic Senate. Senate review.
May 2017	Divisional Senate recommendations received. If supported, submission of pre-proposal to UCOP provost to be forwarded for review as specified. Preparation of full proposal begins as reviewers' comments are made available.
August 2017	Submission of full proposal to campus administration for review by all necessary instances, as outlined in the BRPG and Compendium.
May 2018	Review beyond the campus completed as needed. If supported, approvals in place.
July 2018	Establishment of new school.

Strategic academic planning for FTE

Responsible parties in the campus administration and the Senate will have the lead in this area. No tasks need be recommended by the FAB.

Fundraising

Responsible parties in University Development and Alumni Relations will be able to partner with the director of the data science initiative and its executive and academic committees. Fundraising for the undergraduate data science education program should continue and is in a position to be ramped up in collaboration with the Dean of Undergraduate Studies in L&S. Designing a compelling fundraising pillar of the character recommended in this report necesssary involves strategic input from deans and the Senate. The process should be moved forward on a rapid timescale in connection with current governance paradigms for significant fundraising proposals and the development of the next campaign.[57]

Growth of the undergraduate Data Science Education Program

The DSEP's progress to date has been outlined above. Begun in AY 2015-16, it started at the introductory level with the cross-listed Foundations of Data Science class (CS/Info/Stat C8, total enrollment including fall pilot ~550 students in AY 2015-16) and 12 connector courses. All courses offered in the Data Science Education Program have been approved by the Senate's Committee on Courses of Instruction. The program has been steered to date by the leadership of the Data Science Planning Initiative and has operated from an interim home provided by the L&S Dean of Undergraduate Studies.[58]

---

[57] In addition to partnership with UDAR and existing decanal fundraising staff, data science initiative requires at least one full-time FTE of high-level dedicated fundraising support with significant experience in this sector.

[58] The DSEP (http://data.berkeley.edu/data-science-education-program) has been launched with faculty and departmental contributions in time and in kind, substantial assistance from the L&S Undergraduate Dean's staff and reserves, temporary support from a campus allocation of TAS and administrative funds, and fundraising and donations assisted by UDAR, the Department of Electrical Engineering and

Next steps for the program involve reviewing progress for its launch phase, continuing introductory-level offerings at roughly the same scale for AY 2016-17 (each semester ~500 students in the Foundations course and ~10 connector courses) and increasing scale AY 2017-18, expanding outreach to additional departments, and creating upper-division classes for approval by COCI and pilot offerings in Spring 2017. It is anticipated that proposed major and minor programs can be developed for review by the Senate in AY 2016-17. Depending on approval timelines and availability of resources, students may be able to graduate with data science majors and minors in AY 2017-18. Much detailed work, consultation, and coordination will be involved in these efforts, beyond what is specified here.

To provide essential leadership and an interim organizational basis for these developments as planning moves forward for a new school, the L&S Division of Undergraduate Studies has agreed to provide an interim administrative home for the DSEP. The leadership of the DSEP should be asked to prepare financial and other plans for campus review. We recommend that the next two years of the DSEP be supervised by the data science initiative with the cooperation of the Dean of L&S Undergraduate Division as outlined above.

Strengthening and integration of data science in research and graduate education

Research and graduate education in data science are intrinsically tied to the many places on campus where it unfolds. It is attractive to use the cross-roads functions of the Berkeley Institute for Data Science, in partnership with the data science initiative, to explore where there are ways to build on, connect, and expand the campus's strength in these areas. It is essential that faculty from all units feel strongly welcomed to become part of shaping such efforts of planning and execution, providing both input and collaborative leadership.

In consultation and collaboration with existing campus units and programs, it will be possible to leverage and extend BIDS's campus outreach in research, ideally with the aim of generating network effects (connecting faculty with faculty, as well as with BIDS). This form of work will prototype the efforts of the new school to build and sustain connectivity in the applications of data science in research domains. It will be valuable to engage additional faculty from many domains to explore what kinds of engagements would assist them in pursuing data science research.

It should also be possible to work collaboratively to envision possible both shared and distributed elements in developing new graduate offerings, building on courses (including upper-

---

Computer Sciences, and the College of Letters & Science. The initial allocation of funds for the DSEP covered through Fall 2016. Through in-kind contributions and close stewardship of resources, Spring 2017 instructional costs can be covered with funds carried forward from the initial allocation. The DSEP should be integrated into regular TAS and classroom space planning for AY 2017-18, and fundraising to accelerate its growth should expand. Existing connections to the Library, IS&T, and other campus units can be strengthened as the DSEP is regularized organizationally.

division undergraduate courses in the DSEP) and programs (including existing masters degree curricula, new ones being designed, and Designated Emphases) that have been or are being developed across the campus. This work can productively begin with an informal gathering of faculty and programs involved in graduate-level data science education (masters and PhD level) planned for September 2016 under the shared leadership of the BIDS Education and Training Working Group and the DSEP.

Other efforts

Separate from the data science initiative, but in coordination with it, important elements of university administration and infrastructure provision will wish to continue efforts to accommodate the impact of data-intensive research and teaching across the university. The FAB recognizes efforts underway in the Library, IS&T, ETS, space planning, and elsewhere. While it recognizes financial constraints, it emphatically encourages a continuing emphasis to develop programs and service offerings in this area.

14. Conclusion

The impact of data is being felt across many sectors, carried forward by the pervasive digitization and instrumentation of the natural, engineered, and human worlds. Researchers are collecting staggering amounts of data to refine existing theories and to discover new phenomena, with fields such as genomics, astronomy, neuroscience, particle physics, and earth sciences in the vanguard, and many other areas in the natural, social, and cognitive sciences, the humanities, and the engineering disciplines in close parallel. Advances in artificial intelligence are fueled by huge new datasets that are powering machine learning and deep learning applications. In technology and commerce, the proliferation of mobile devices and networked communities has meant that humans have become both consumers and providers of data, and businesses and governments have begun to create new services and platforms that use data in order to be increasingly responsive to evolving concepts and individual needs. Organizations are investing in generating actionable guidance from their own data, looking to chart paths through a landscape that seems to shift more dynamically and rapidly than before.

The momentum behind data science is nationally established, with most leading academic institutions launching visible, resourced initiatives, federal agencies making it a strategic priority, and powerful industry focus apparent. At present Berkeley holds a unique position of leadership in faculty and research strength. Our next steps need to be strong and strategic, connected and flexible, tuned to the distinctive promise that the world's best public university holds. We see a once-in-a-generation opening for Berkeley to define the global terms of the field in data science.

The Faculty Advisory Board believes that a vision that is simultaneously deep, broad, and rich can both draw Berkeley together in data science and make a distinctive contribution to the world. It involves laying the foundations of the field and pushing its conceptual frontiers, applying established or emerging technologies and techniques to the wide range of areas or domains, and studying the implications of the explosion of data and analysis for ethics, policy, society, and human knowledge. Building on Berkeley's existing strengths and our experience with campus-crossing organizational forms, we have articulated four expectations to be met by the university's efforts in this domain. These are maintaining and strengthening faculty excellence, offering outstanding education at scale, developing strategic organizational capacities for the data science area, and realizing Berkeley's vision for the field as a whole. Our recommendations for realizing these goals are threefold: to create a school for computation and data science, instantiated in one of two feasible organizational forms, with a mandate to realize a constructive culture of border-crossing engagement; to invest in a strategic academic plan for faculty positions on the scale of 20-25 FTE strength, allocated across multiple units and distributed over ten years; and to create a pillar of fundraising across the university that supports both the school and other academic units, with deliberate attention to new endowed faculty positions, among other things.

This plan is a distinctively Berkeley approach to the challenge presented by the still-shifting landscape of computation, connectivity, and analytics. It continues the trajectory that has brought our faculty to this point, moving us into a new phase of experimentation under the guidance of a deliberate strategy. It is an ambitious proposal. Our hope to deliver on depth, breadth, and richness in data science is in line with the multiple dimensions of Berkeley's excellence and public mission. Realizing that aim is an institutional challenge, for certain. Our organizational paradigm is also challenging. Where other universities have sometimes made choices—building a compact academic core for vertical strength, or creating an institute stressing horizontal connectivity and breadth—we see it as appropriate for Berkeley to do both. Success will depend on principled design, experimentation, and close attention to the on-the-ground reality of connections between disciplines. It will also depend on committed leadership that moves forward the institutional arrangements for the culture and resources necessary to make the plan work, including substantial new efforts in philanthropy and revenue generation.

Now is not an easy time for Berkeley to contemplate an effort of this character. We have assessed the situational challenges with the information at our disposal. We recommend that the university choose to move from contemplation to action, as we see a data science initiative of this character as a responsible investment in Berkeley's future, given the assets we have, the opportunities we can see, and the transitional costs we believe we can manage. The university will be investing in the capacity to push ahead in this domain scientifically and nurture new intellectual areas, to provide faculty strength to meet a transformational need across many domains, to build robust graduate and professional programs, to support an undergraduate curriculum that instantiates both core strength and connectivity, and to deepen our ability to draw in philanthropic partners from a part of the landscape to which Berkeley has only begun to reach out. Underlying each of these measures is an investment in the strategic capacity to move effectively in this area, now and as it continues unfolding.

The university's usual incremental processes are unlikely to bring us to the outcomes we hope for. At the same time, any strategic processes we create should be principled, realistic, and respectful of the elements of shared governance that have made Berkeley strong. The FAB believes that the next stages of the university's efforts in data science can themselves be a means to bring us closer to realizing our ambitions. We urge that the university carry out the processes that will follow in a way that is designed to strengthen the outcome. Faculty leadership and a focus on our strategic intellectual future are at the center of our recommendations. Openness and fair-mindedness are key, making process something more than a set of procedural steps to walk through.

These next steps matter, we believe, for Berkeley's future, both for computation and data and for our broader strategic positioning. As significant as the reorganization of biology was in the 1980s, we see strong parallels in the present moment, making the effort important enough to take on even in a difficult times. Our goals are ambitious, as goals rightly have been at previous moments of strategic academic investment at Berkeley. When we design for the future, it is guaranteed that things will not happen exactly as we plan. Weighing the risks of action against the risks of inaction, we think the most responsible approach is to proceed based on grounded design principles and an openness to experimentation, all of us sharing responsibility for the good of the whole.

A1. FAB Charge, Constitution, and Membership

Charge to the FAB

The Faculty Advisory Board of the Data Science Planning Initiative is charged with developing an integrated strategy for Berkeley's global leadership in data science. In its advisory role to the Chancellor and Provost, it will chart paths of institutional development that can support a comprehensive initiative in data science, drawing out and advancing Berkeley's distinctive strengths throughout this domain. It is expected that the DSPI will encourage the confluence of data science activities across the full range of our research and teaching, including the core specialty areas of data science, its use in research domains across campus, and its broader societal and normative entanglements. In developing its vision, the FAB will assess our university-wide strengths, gaps, connections, and opportunities in data science and address questions of coordination, faculty hiring, resourcing, fundraising strategy, and organizational forms. In order to prepare for significant investment in this area, the DSPI will engage the campus community in the formative stages of its process. By the end of Spring semester 2016, the FAB will document its strategic vision in the form of a white paper for the Chancellor and Provost, together with whatever interim reports and supporting materials it finds useful to provide.

Constitution

The 15-member Faculty Advisory Board was formed by the Chancellor and Provost to do exploratory and advisory work in advance of formal processes that might follow. It was constituted as an administrative committee composed of faculty members in their personal roles. Its members were asked to draw on their expertise and to reach out broadly for perspectives and information, rather than to view themselves as representatives of departments, colleges, or schools. In its constitution it was positioned to seek out input and exchange with elements of campus academic administration (chairs and deans) and with the Senate while operating on its own standing. It drew upon diverse experience in the institutional Senate (current and past roles include BIR, CEP, CAPRA, COCI, DIVCO, Graduate Council, P&T, and Chair of the Berkeley Division) but was not formed as a body whose constitution would be part of the mechanisms of shared governance at this stage. Its recommendations are advisory.

The Data Science Planning Initiative, as part of which the FAB was constituted, operated from 2015-2016. Its creation was announced to the campus in a Cal Message of November 3, 2015,[59] and the FAB was constituted shortly afterward.

[59] https://calmessages.berkeley.edu/archives/message/39522

FAB membership

Cathryn Carson	History, DSPI FAB chair
Lisa García Bedolla	Graduate School of Education and Political Science
Francesco Borrelli *	Mechanical Engineering
Ron Cohen	Chemistry and Earth & Planetary Sciences
David Culler	Electrical Engineering & Computer Sciences
Rosemary Gillespie	Environmental Science, Policy, & Management
Sol Hsiang	Goldman School of Public Policy
Bob Jacobsen	Physics and L&S Undergraduate Division (Dean)
Michael Jordan	Statistics (Chair) and Electrical Engineering & Computer Sciences
Susan Marqusee	Molecular & Cell Biology and QB3 (Director)
Anno Saxenian *	School of Information (Dean)
Jas Sekhon	Political Science and Statistics
Chris Shannon	Economics and Mathematics
Ion Stoica	Electrical Engineering & Computer Sciences
Bin Yu	Statistics and Electrical Engineering & Computer Sciences

Members indicated with a * did not sign the FAB report.

A2. Additional Considerations in Organizational Design

Organizational options reviewed

There is a finite number of ways to provide core organizational support for the data science initiative. We list below options that have been mentioned at various stages in the process. We have touched upon them at greater or lesser length within the FAB but moved to focus on the few options that are responsive to the goals. If the question is asked, “Why not this particular option?” we are happy to make ourselves available as a body to discuss.

1. Create a college
2. Create a school
3. Transform an existing school
4. Create a new division in L&S
5. Transform an existing division of L&S
6. Create two divisions within Engineering
7. Create a division that is associated both with L&S and Engineering
8. Create a program within a school or department
9. Create a program that is spread across a collection of departments
10. Create a undergraduate teaching program
11. Create a research unit that can hold fractions of faculty FTE
12. Create a New Initiative Center
13. Create a graduate group
14. Create an augmented graduate group
15. Create a Designated Emphasis
16. Create an ORU (alternatively, a center, an institute, or a lab)

Options 13-16 may be useful vehicles for particular purposes, even as they cannot meet the goals for the core effort as a whole

Administrative consequences for degree programs of a divisional structure, particularly in the cross-college model

One may inquire about the tangible effects on degree programs of being associated with a decanal division, particularly if the concept of a decanal division is not a familiar one. Regardless of divisional recognition, admissions and degree-granting authority would continue to reside with the colleges if academic units are formed into a school in this model. The notion of a degree awarded by a college being delivered by a faculty in a decanal unit distinct from the college exists today and needs no invention. The CoE faculty in EECS has delivered the L&S Computer Science Bachelors degree for more than four decades. Students are admitted to the College of L&S and declare the major. Similar situations are present with Public Health, Business, Operations Research and Management Science, and others. Substantial (cross-divisional in L&S) staff organizations exist, especially for admissions, advising, and degree process, but also for fundraising, course management, and other functions; as a division these are naturally available to support the school that is formed largely by bringing together units in the colleges. In comparison, being only an “administered program” (as Computer Science, Cognitive Science, and ORMS were for a significant period) places units in a second-rate status and is arguably unfair to students within those programs. Divisional status within two colleges allows cross-college programs of the sort we would like to encourage to be treated by each as a peer of traditional programs, with the sense of ownership and identification. New programs, such as bachelors and masters degrees in data science, are natural to make available to students in either college (or in schools) and can be delivered efficiently through the joint effort of the core across the school.

A3. Additional Considerations on Process to Follow

The data science initiative will require commitments of administrative and faculty time to execute its tasks successfully. To bring the campus to effective outcomes with minimal disruption, processes involved in preparing organizational proposals, assessing resources, and developing transition plans for a new school need to be executed carefully and thoughtfully. For the fundraising effort, dedicated development staff will be required.

The FAB was asked to provide only high-level guidance on the transitional effort required. To scope the work, we offer a rough estimate of staff commitments required directly within the DSI:

1. Administrative Manager / Officer position reporting to Director, 50%
2. Senior Director of Development position reporting to Director, 100%
3. Existing DSEP staff roles continued through June 30, 2018

This staffing model is approximate and only for scoping; the faculty director of the data science initiative should be asked for a concrete plan. Faculty time is also a key resource to allocate, both for the director and potentially for other faculty in supporting roles. We believe the university will need to invest in making faculty time available through course relief and other standard means. We suggest that it may be appropriate to identify service to the initiative for its duration as campus-level service, in order to make it possible for faculty to devote service to this effort in lieu of departmental-level service as needed.

Our ballpark estimate of cost for a two-year organizational / transitional effort (AY 2016-17 and AY 2017-18) is $2.5M. To enable the transitional process will require drawing on some campus investment, while recognizing that this central allocation cannot be large and may in fact require raising philanthropic funds to cover part of the cost. We believe that this nontraditional approach may find some support among a particular class of donors who care about institutional transformation at the university and are familiar with VC practices.

Carrying out this work will also require staff effort on the part of offices outside the data science initiative, including the central administration, deans' offices, departmental staff, and University Development and Alumni Relations. We encourage careful consideration of staff workload.