Summary of the 2016 Report by the Faculty Advisory Board of the Data Science Planning Initiative

Prepared: 19 August 2016
Cathryn Carson, FAB Chair

Contents

In summer 2015, the Chancellor and EVCP established the Data Science Planning Initiative, which included a Faculty Advisory Board (FAB) charged with writing a report on Data Science at Berkeley. This document is a summary of and guide to that report.

A. Rationale for action: Why Berkeley, why now

New data collection technologies and computational and inferential approaches to data analysis are transforming many fields of human action, including the production of knowledge itself. Faculty across our campus see their fields shifting, while our students are placing fast-rising demands on the university to educate them for a world in which data-intensive, computationally enabled work opens up new conceptual and practical vistas. At Berkeley this ground-up transformation has been unfolding in a truly distinctive way, with many of the world's leading scholars located here. Inside our university, data science has reached an intensity and breadth, as well as a state of existential challenge at its core, that requires us to take strategic, large-scale, and rapid action.

Right now, as Part I of the Faculty Advisory Board (FAB) report outlines, Berkeley has substantial strengths in data science, starting from a world-leading collection of faculty along the axis from statistics to computing, situated in exceptionally strong programs that are conspicuously collaborative between themselves and across their boundaries to application domains. In our broad-based excellence Berkeley has strength across many fields in which faculty use large-scale data collection and analytic methodologies to rigorously address research questions, from the social sciences and humanities to engineering and the sciences

and across the professional schools. We have a distinctive commitment to openness and a public mission that already underlies many of our engagements with data science, and we have the powerful intellectual reach and traditions of free discussion that let us critically examine how data science shapes human societies and lives. We have institutional innovation in masters and doctoral education and in research-supporting infrastructure for this area. Finally, we have the unique willingness of our faculty to engage in ground-up collaboration to meet campus-crossing needs, as in the Berkeley Institute for Data Science and Berkeley's rapidly expanding data science undergraduate education program.

These assets are here, yet they are suboptimally organized, left without fully adequate support, and unsustainably set up for the future. To keep them, grow them, and reap the fullest benefit of having them, we need strategic action at significant scale very soon. Faculty at the core of this area see its rapid spread across sectors of society and through the higher education landscape. They have begun making their choices about whether Berkeley can adapt before the moment is lost. If we do not take significant near-term action, we undercut some of our strongest programs whose trajectories are at this moment bringing them closer together, and we miss the chance to equip the university to move adequately into the data-intensive future that so many of our faculty see.

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. If we take this opportunity, we see a once-in-a-generation opening for Berkeley to define the global terms of the field in data science. 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. It has, moreover, strong practices of collaborative faculty-led, administratively supported effort in this area that the FAB can hope to continue.

B. Recommendations

The FAB believes Berkeley needs to move rapidly to create a powerful, sustainable platform for the varied aspects of data science to thrive. There is no existing model to copy, so we have to approach the problem with inventiveness, rigor, and principled thinking. We must build both for strength now and for adaptability over time. Ultimately we must rest our approach upon the practices of faculty judgment and shared governance that have proven key to Berkeley's excellence, while fitting them for this new cross-cutting strategic demand.

In the face of the need and the timescale for action, the FAB's report identifies the goals we believe the the university must meet in data science: maintaining and strengthening our faculty excellence, offering outstanding education at scale, developing strategic organizational capacities for the data science area, and realizing Berkeley's deep, broad, and rich vision for the

field as a whole. Part II of the report details the rationales that lead us to recommend three interdependent actions to be taken together. We summarize those recommendations here.

1. Organizational form: Core and connections

After reviewing all viable options that we believe would find the assent of a significant body of faculty, we see one organizational pathway to Berkeley's excellence in data science that satisfies the goals on the timescale we need. **It is a structure that combines a) a core academic unit of deep intellectual strength, constructive outward-facing culture, and decanal stature with b) broad, strong bi-directional connections that bring value to the core and to domains campus-wide.** With the strong collaborative connections already in place and the interest expressed in a prospect of this sort by faculty in programs including Electrical Engineering and Computer Sciences, Statistics, the School of Information, several interdisciplinary programs (computational biology, new media, and others), and possibly emergent clusters of faculty who may choose to come together, it is the FAB's strong sense that a world-leading unit can be built. The interests and needs of each of these groups of faculty have shaped the FAB's reflections on the course we recommend following. This move represents a faculty-driven co-location of programs with strong synergies that currently sit in disparate decanal units, with potential reconfigurations that can happen if they are brought together. The size of this grouping is potentially on the scale of our medium-sized to large decanal units (larger schools, colleges, and L&S divisions).

We recommend that the campus embark on the processes that can lead to the creation of this decanal unit in the shape of a new school centered on computation and data science. Having considered multiple options as detailed in the report, we recommend that the goal be to form this school administratively as a division of two existing colleges (Engineering and Letters & Science). It could also exist as a free-standing school. An explication of the rationale for this construction can be found in Section 6 of the report. We see a strong core with decanal stature as essential to exercising strategic capabilities in academic personnel planning, resource allocation, and curriculum planning for key parts of this burgeoning area that serve the campus at large. It needs independent decanal strength, moreover, to have a voice in shaping the campus's direction and to pursue systematic fundraising. 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.

The culture and leadership of the school need to integrate and deepen the long-term practices of engagement across boundaries already on view in our strong academic units in this area. It needs to reward the collaborative, constructive behaviors incentivized by the nature of research challenges in data science itself. Data science crosses intellectual boundaries with a huge array of capacities to address problems in different domains. Berkeley's successes in this area so far have derived from our strong collaborative culture already in place and our willingness to experiment with appropriate forms domain by domain. **The school may find itself the home of one or more new academic subunits of some form around aspects of**

computation and data bridging strongly to fields outside it. Within the school the Berkeley Institute for Data Science could appropriately flourish as a crossroads and a platform for application-oriented collaborations, open/reproducible practices, and targeted service to science domains situated outside its boundaries.

It goes without saying that this school will not "own" data science everywhere across the university. Our diverse landscape of data-related research and education is a source of strength for the Berkeley campus. In the FAB's 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.

Faculty-led processes for forming the school, following an opt-in approach available to both existing academic units and individual faculty, fit into existing mechanisms defined by the Academic Senate and administration and respect individual faculty autonomy throughout. It will be critical to create a fair environment for principled discussion among academic units with significantly different histories, profiles, and sizes. Faculty not yet organized into programs need ways to participate and, as appropriate, join, and internal boundaries within the school among departments and programs are open to being reconfigured. Preparing proposals to constitute a school and securing Senate and administration approval can take two years. The FAB believes that the most important thing Berkeley can do organizationally is to give the signal to its faculty that it intends to support action toward forming a school.

The FAB recognizes that this recommendation entails a substantial effort. We would not recommend it if we did not think it was essential and possible. 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.

2. Faculty FTE: Campus-wide surge and strategic foci

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. **We recommend that the campus invest in an influx of data science faculty positions. 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. In Section 8 of our report, the FAB carries out a modeling exercise, based on simple, adjustable assumptions about developments in domains both inside and outside the core. **Our conclusion is that a mix of positions on the order of 20-25 FTE strength over ten years is necessary.** While this may seem large, even in a peak year of, say, 5 new FTE, the surge would amount to a small fraction of total faculty hiring campus-wide.

It is not for the FAB to prescribe how this process should be implemented. Allocation of faculty FTE at Berkeley is steered by processes of shared governance. Between the administration and the Senate, the campus must reach a shared understanding of a strategy for academic planning for data science.

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. Each of the opportunities for strategic foci examined in Section 9 of the report offers distinctive features around modes of institutionalization inside or outside the new school. 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 pillar and revenue generation

As outlined in Section 10 of the report, computation and data analytics offer significant opportunities to develop Berkeley's capacity to draw upon philanthropic resources. **We recommend moving 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.** 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. 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.

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. This fundraising 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.

There are significant opportunities for revenue generation in data science at Berkeley, both inside and outside the new school. Section 11 of the FAB report indicates some key possibilities apparent to the FAB, beyond the opportunities currently being exploited.

C. Situational challenges and next steps

The university's usual incremental processes are unlikely to bring us to the outcomes we hope for in data science. Thus Part III of the FAB report offers observations on possible next steps. Any strategic processes the campus creates should be principled, realistic, and respectful of the elements of shared governance that have made Berkeley strong. We will need to come to terms with situational challenges (at Berkeley, at present) in order to realize the report's recommendations. Section 12 of the report outlines challenges relating to the severe financial situation in which the university finds itself, the novelty of the efforts these proposals will require, and the overall challenges of designing for interdisciplinarity, flexibility, and adaptivity.

The FAB strongly believes that a realistic assessment of our situational challenges still points us to proceed. We offer a schematic outline of processes in the form of a next-step data science initiative in Section 13 of the report. In full consciousness of the demands on the campus at present, the FAB believes that we must invest our effort, now, in a series of significant next steps in line with the three recommendations above. Along with the intellectual and institutional motivations, the possibilities opened up for significant philanthropy across the campus and for strong, scalable revenue generation absolutely must not be left untapped. We believe, finally, that data science offers a unique opportunity to bring Berkeley's faculty together around a shared effort that has the potential to bring benefit to a large number of domains.

D. The Faculty Advisory Board

The Faculty Advisory Board is composed of 15 faculty from across the university appointed by the Chancellor and Provost as an administrative committee following familiar campus procedures. Its members were asked to do exploratory and advisory work in advance of formal processes that might follow. Its charge and constitution are described in an appendix to the report. 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.

The Faculty Advisory Board is one of the principal components of the Data Science Planning Initiative constituted in summer 2015. More information is on its website at data.berkeley.edu. The FAB reports to the Chancellor and Provost; the DSPI co-directors serve on it ex officio.

The FAB worked for 8 months in 2015-16 to develop its report. It achieved consensus on the outlines of the solution presented here early in spring 2016 and was able to submit an unsolicited memo to the campus's strategic realignment exercise (from which the DSPI was

separate). This summary is derived from the FAB report submitted in summer 2016, which drew on the input of all members of the FAB while representing the views of those members who signed the final report report.

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.