



November 14, 2016

CAROL CHRIST
Interim Executive Vice Chancellor and Provost

Subject: Report from the Faculty Advisory Board of the Data Science Planning Initiative

Dear Carol,

The Divisional Council (DIVCO) discussed the Faculty Advisory Board (FAB) report of the Data Science Planning Initiative at its meetings on October 24, 2016, and November 7, 2016. DIVCO's discussion was informed by the report of the ad hoc working group on the FAB report. The working group was composed of the chairs of the committees on Academic Planning and Resource Allocation (CAPRA); Budget and Interdepartmental Relations (BIR); Educational Policy (CEP); Graduate Council (GC); and Undergraduate Council (UGC). These chairs discussed the FAB report with their respective committees, and these committee discussions formed the basis of the working group's deliberations. The Divisional Vice Chair chaired the working group. The working group report is appended.

DIVCO supports a rapid and aggressive move into the intellectual space of Data Science by the Berkeley campus. Any new structure must be externally visible, properly resourced, and fiscally responsible. DIVCO also believes that this new structure should be a division.

DIVCO recognizes the potentially significant cost of creating a new division, and this consideration figured prominently in our discussion. But there was wide agreement that lower-cost alternatives, e.g., a New Initiative Center or an augmented graduate group, have an uneven track record of success. There is a significant risk that the momentum behind Data Science on campus will dissipate absent the structure and resources of a division. Looking to the longer run, DIVCO encourages the administration to consider organizational simplification and consolidation that might reduce the number of decanal units on campus.

In addition to campus resources, ultimate success will also require substantial philanthropic support, especially in connection with externally funded faculty.

The governance structure of a new division is a critical decision, and several options are currently being discussed across campus. These include the bi-college option proposed in the FAB report in which the division would be part of both the Colleges of Engineering and Letters and Sciences. Another option is a stand-alone school, possibly entailing a reconfiguration of the Information School. Still another structure is a new division within the College of Engineering.

DIVCO is agnostic about the appropriate governance structure at this point in the process. DIVCO does however strongly believe that the decision about structure should be guided by several important considerations. First, the ultimate structure should reflect the bottom-up desires of the faculty involved. Second, the structure should support the important goal of a data science initiative that is deep, broad, and rich. Finally, the governance structure should be open and forward-looking, one that allows for the evolving nature of the fields involved.

Should the administration decide to establish a data science division, DIVCO recommends that the next steps should be a timely decision regarding the governance structure followed by a detailed proposal for the new division which would subsequently be reviewed by the Senate and others as specified by the Review Process Guide. Noting the wide range of views across campus about the appropriate divisional structure, or even whether there should be a division, DIVCO discussed the merits of convening a retreat charged with resolving this issue and, ideally, finding a consensus. A respected member of the faculty would be tasked by you with bringing the relevant parties together in a day-long retreat with the goal of reaching a decision on the governance structure. This would be done on an expedited basis, and the retreat would ideally be held over the winter break.

While I want to bring this idea to your attention, DIVCO itself was divided on its merits. Some members see this as a way of expanding the circle of consultation and possibly yielding a widely-shared if not unanimous recommendation about the structure. Such a shared recommendation would be an enormous plus in moving forward. Others see this as largely redundant given the wide-ranging input you have gotten and are getting to the FAB report from the Senate, Deans, several departments, and other parties as well as from the bottom-up efforts that preceded the FAB. There was also some skepticism about the prospects of reaching a widely-shared recommendation. Whether you decide to convene a retreat or not, DIVCO believes that a timely decision on structure is important.

After the governance structure is decided, DIVCO recommends that a faculty lead be named possibly along with an advisory board. The lead will be responsible for working with the relevant departments and units to develop a detailed plan for the new division. The FAB report describes the goals and a timeline for this process. In addition to those goals, the Senate working group emphasizes that any actual proposal for a new administrative unit requires a detailed study of the cost of its creation and operation -- with particular attention paid to new administrators. The proposal would then be reviewed by the Senate and administration as per the Review Process Guide and Compendium.

The Senate looks forward to working with the administration and campus as a whole to ensure Berkeley's excellence and leadership in Data Science.

Sincerely,

A handwritten signature in black ink, appearing to read 'R. Powell', with a horizontal line underneath.

Robert Powell
Chair, Berkeley Division of the Academic Senate
Professor of Political Science

Encl.

Cc: Lisa Alvarez-Cohen, Vice Chair, Berkeley Division of the Academic Senate
Sanjay Govindjee, Chair, Committee on Academic Planning and Resource Allocation
Stuart Russell, Chair, Committee on Budget and Interdepartmental Relations
Max Auffhammer, Chair, Committee on Educational Policy
Whendee Silver, Chair, Graduate Council (GC)
Mark Stacey, Chair, Undergraduate Council
Sumei Quiggle, Associate Director staffing Graduate Council and Undergraduate
Deborah Dobin, Senate Analyst, Committee on Academic Planning and Resource Allocation
Linda Corley, Senate Analyst, Committee on Educational Policy

Report of the Ad Hoc DIVCO working Group on Data Science

November 6, 2016

Background:

After receiving the Faculty Advisory Board (FAB) report of the Data Science Planning Initiative from EVCP Christ, Senate chair Powell formed an ad hoc working group to report back to DIVCO. The members of the working group were Lisa Alvarez-Cohen (working group chair and Division vice chair), Bob Powell (division chair), Mark Stacey (UGC chair), Max Auffhammer (CEP chair), Whendee Silver (GC chair), Sanjay Govindjee (CAPRA chair) and Stuart Russell (BIR chair).

The working group was charged with synthesizing the discussions of the Senate committees. This synthesis report was the basis for DIVCO's discussion of the FAB report at its meetings on October 24, 2016, and November 7, 2016.

Each of the committees met one or more times to consider and discuss the four documents forwarded by Division chair Powell on August 31: 1) the DSPI charge letter of November 30, 2015; 2) the Summary of the 2016 Report by the Faculty Advisory Board (FAB) of the Data Science Planning Initiative (DSPI) prepared by FAB Chair Cathryn Carson on August 19, 2016; 3) the full DSPI report dated 2016; and 4) the memo providing an alternate perspective by Dean Anno Saxenian of the School of Information and co-chair of the FAB dated August 11, 2016.

The working group's synthesis of the committees' discussions follows:

Summary of Committees' Discussions:

The committees considered first the overarching vision: Should Berkeley be building a new research and teaching initiative around data science? Second, they also evaluated, in general terms, the appropriate structure for a potential data science initiative: Should this be a new school or another structure? Third, they evaluated the general financial model: Where should the funds for this initiative come from? The discussions of the committees have been synthesized and are structured into four areas: Intellectual Merit, Proposed Administrative Structure, Educational Component, and Resources. The Committees agreed that the FAB report focuses on institutional structures, and only briefly touches on undergraduate and graduate education and curricula. Reference is made in the FAB report to the previously released DSERAT report (from early 2015), but the comments here are based strictly on the FAB Report, as noted below.

Intellectual Merit:

The FAB Report outlines an institutional approach to Data Science that involves a concerted and aggressive investment to establish Berkeley as the intellectual leader in this field. The committees expressed enthusiasm for strengthening the research and teaching of data science on the UC Berkeley campus and agree with the report that strong support should be provided for a bold and rapid move into this

intellectual space. Not only is there significant student demand for the field, we also believe there is an important intellectual pursuit in inferential and computational analysis of data that Berkeley can and should continue to lead. Therefore, discussion on the **intellectual merit** of a significant effort in data sciences was well supported by the Senate. A few issues were raised, as discussed below.

- a) **“Depth, Breadth and Richness”**. Any effort of this scale, of course, creates concerns or uncertainties; and the committees highlighted several. These should not be interpreted as opposition to the vision of pursuing Data Science aggressively at Berkeley; rather, they are intended to help shape the conversation about the details of the initiative. *“Depth, Breadth and Richness”* in Data Science are the three themes that the FAB report develops to describe the broad intellectual scope of the new unit. An investment in Depth is obviously central to any Data Science effort, but the other two components were points of focus in committee discussions. It is important to note that we emphatically *do not regard the proposed new Division’s scope as Data Science*; it is considerably broader than even the most liberal interpretation of that term, and a very large number of faculty in the associated units do not consider themselves data scientists. The emergence of Data Science as a campus-wide theme does, however, provide a strong impetus for a reorganization.

Principle 1:¹ Berkeley should explore a highly visible and well defined effort in the space of data science. We suggest that the campus flesh out Berkeley’s broader vision for the future of this field and the big questions that this effort would attempt to answer, instead of focusing the effort on tools and methods.

Proposed Administrative Structure:

The FAB proposes an administrative structure for the new unit that on face value appears complex and perhaps costly. Unfortunately the FAB does not discuss in detail the constraints under which this proposal was developed and thus it is difficult to judge if it is the right structure or if there is a less costly alternative. The FAB Report’s main proposal is an organizational realignment forming a new academic unit that is envisioned as a Division within both Engineering and Letters & Science, whose initial members are likely to include the current EECS and Statistics faculty and may include the iSchool, IEOR, Computational Biology, Demography, and possibly others. However, the justification of a division or school, in terms of the University’s research and education missions and the broad landscape of intellectual inquiry, is largely missing. Providing a clear justification is important given that such reorganization will have major implications for many years. We believe that the justification could be built on the interlocking foundations of computation, information, communication, statistical modeling and inference, and optimal decisions, rather than around technical considerations such as increases in

¹ Principles are enumerated so that they can be referenced more easily. The numbers do not indicate anything about relative importance.

processing power or the pervasive availability of large amounts of data. The proposed structure appears in part to be designed to avoid the major negative impact on the College of Engineering that might result from EECS (or some large fraction of its faculty) moving into a separate, stand-alone School. Under the proposal, Engineering will have to decide whether to form two Divisions, or have the new Division and several other Departments. In the latter case, it is unclear whether there would still be a dean for the collection of departments outside of the new Division.

The committees would like to lend support for rapid and aggressive action towards an institutional structure that will allow the field of Data Science to flourish at Berkeley, and for Berkeley to maintain its international leadership in the field. This leadership has emerged largely from Statistics and Computer Science faculty.

While some committees believe a new structure like the one described in the FAB is likely to best enable the campus to move quickly and effectively to maintain and grow Berkeley's presence in the field, this was not a universally held belief. The question was raised as to whether the Information School could be reconfigured to serve the identified need, rebranding itself as a new School for Data Science (or other name) that could rapidly be expanded to fulfill the data science mission by inviting faculty in other units to become affiliated with zero or partial FTEs. The structure of a school enables the possibility of new departments within the school -- for example a liberal arts oriented department associated with data science, its applications, as well as its social implications. Though of course such a department could also arise in one of the existing L&S Divisions and affiliate with whatever new unit is developed.

A possible lower cost structure would be a New Initiative Center (NIC) that could serve as the home of the undergraduate major and graduate programs while faculty retained their homes in disciplinary departments. It could have a faculty director and an executive committee and receive a direct allocation of FTE and TAS funds. This would certainly be much lower cost alternative than a new school and would achieve the goals of high visibility and the potential for direct resource allocation. Current faculty would retain their identities within their current departments while affiliating with the Center and new faculty would choose their home departments leading to both depth and breadth in hiring within data science. With 80 faculty behind the initiative it is unlikely to face the issues that our less successful NICs have faced on the campus. It was felt that it is not necessary to have a major structure like a new school and its full apparatus to attract large donations and that we should be cautious of the "*if we build it they will come*" mentality. Donations depend on a clearly articulate idea and vision, as demonstrated by the recent Zuckerberg-Chan \$3 billion donation for a *research center*.

For any of these 3 structures or others that are finally adopted, a fundraiser could be assigned by UDAR to work with the Dean, Director or Executive Committee.

Principle 2: As part of the final proposal, the constraints associated with the final recommendation need to be transparently discussed. Without that, it will be very hard to make an informed recommendation. The school within two colleges is seen as an expensive and likely very inefficient structure; its main advantage is that it could be viewed as a solid dramatic change that could be highly visible. A standalone school has the same characteristics but would likely be able to operate more efficiently. The NIC option is the most cost effective but could be viewed as less dramatic; for the purposes of fundraising, however, this perhaps will matter less than some would argue -- ideas and commitment are what will count more. Finally, it is imperative that any planning-team includes strong representation from non-Computer Science and non-Statistics faculty (with a strong tie to data science from the applications side and the social implications side).

Additional discussions on the proposed structure were lengthy and can be summarized as follows:

- a) **Proposed Academic Home.** Many committee members were skeptical of the proposed split of the unit between two colleges and the administrative challenges this poses, and there is no precedent for this. Members further did not understand why a significantly expanded School of Information was not the natural home for this effort. Members acknowledged that the report outlined several possible structures, but did not fully concur with the proposed solution. This was partially caused by some concern about what this would do to the EECS and Statistics Departments, where it is possible that a significant share of the faculty would move their FTEs from the departments into the new unit.
- b) **Questions to be addressed on structure.** What is the nature of the connection of the new Data Science unit to other units on campus? What is the place of faculty who are interdisciplinary or disciplinary data scientists who do not naturally fit into one of the initially proposed units within computation and data (a designation used in the FAB report) How should additional structures such as graduate groups, centers, institutes, joint majors, and designated emphases be used to accommodate the variety of connections that exist? Will new units join as a whole or will individual faculty members leave an outside department to join the new unit? In the latter case, could they move (some portion of) their FTE to the new unit, without an obvious home department? Under campus policy, any faculty member can join any department that votes to accept them. In this case, who would vote?
- c) **Relations to other Department.** Another important issue is whether the new Data Science unit should include Departments of Computational X (using the terminology of the report)—for example, Computational Biology, Computational Climate Science, or Computational Social Science. How would such departments relate to their non-Computational X

counterparts? In the committees' view, it would be preferable that activities related to Computational X be part of unit X, as we believe strongly that academic units should be organized around fields rather than methods-- a situation quite similar to how applied mathematics is currently handled on campus. However, we acknowledge that in some cases X may not initially recognize Computational X as an intellectually valid approach, in which case the new Data Science unit may provide a nursery of sorts for new areas of activity.

Principal 3: The campus needs to give some careful thought to the most beneficial structure for this new initiative. The campus currently has a proliferation of units with associated Deans, staffing and administrative bureaucracy. If the administration decides that a new standalone structure is best, the committees suggest that a possible implementation task force seriously rethink the academic home/structure for the effort.

- a) **Scale of Effort:** Committee members had varied views as to the proposed scale of the effort. At the minimum it appears that at least 80 faculty are fully behind the initiative (in round numbers 60 from the Division of Computer Science in the EECS department, and 20 from the department of Statistics). This constitutes a large core of faculty who are committed to this path for their disciplines. We acknowledge also the vote of the EECS faculty to stay as one unit thus giving a +40 in some sense to the size of the driving force. With the proposed new structure, there was serious concern about the cannibalization of other campus efforts (e.g. FTE) in the name of data science. Further, some members were sympathetic to a smaller effort in this space at the scale of QB3.

Principal 4: If campus goes ahead with a significant effort in this space, the committee suggests that this effort be mostly funded through philanthropy and other sources of external funding. In the case of faculty FTE, this would require a specification of a tax rate of externally raised FTE for department target FTE numbers.

- b) **Integration of Data Sciences:** While the report discusses the relevance and possible integration of other disciplines/efforts into the new structure, committee members had questions about the consequences of this strategy. If faculty can freely move their FTE into the new "School", this might negatively affect programs that are net losers of FTE. In a worst-case scenario, this could affect the ability of departments to mount their core curricula in the disciplines most affected (e.g. statistics and computing). Further, the committee does not understand how faculty in the non-core disciplines would teach in the data sciences effort. Would social scientists teach social science in this major and one student could obtain a e.g. DS Economics undergraduate degree, while another might opt for a regular Economics degree? Or would these faculty teach connector

courses to the “traditional majors”. What would be the educational integration in the graduate curriculum? Would there be a Designated Emphasis? What would MOUs between faculty and other departments look like in terms of teaching obligations?

Principle 5: If campus goes ahead with a significant effort in Data Science, the integration of faculty should be more carefully thought out taking into consideration the consequences for former home department “curricula”. Teaching obligations should be defined via written MOUs. The relationship of a DS undergraduate program and existing majors and graduate programs should be carefully thought through and fleshed out.

- c) **Richness and Breadth in Data Science.** The committees are supportive of the FAB report's inclusion of the “richness” element of the initiative and the tight integration of data science with its application through the “breadth” element. At the same time, we don't believe the report fully recognizes what would be required to successfully pursue these themes. To be successful in this area, any proposed structure must include social scientists who think about what makes data good, fair and useful, as well as those focused on the Depth theme, in the institutional structure. This means that people with expertise in the social sciences are needed to make computation and data science better, not just that social scientists use data. The minority report from the School of Information was deemed indicative of a general weakness in the structure to incorporate social scientists into the DSPI in a meaningful way.

Principle 6: The incorporation of disciplinary expertise (social and natural sciences, other branches of engineering) into the new structure should be pursued with a focus on the development of the data sciences, not necessarily simply the application of data sciences to other disciplines. An institutional structure built around data science should be centered on pushing the field of data science forwards intellectually, not proselytizing its existence and use. Clarifying and preserving this distinction between the development of the field and the applications of Data Science on campus may also help in managing the impact on social science and other disciplines/departments.

- d) **External Impacts of New Organizational Structures.** With the expansive vision for an institutional structure contained in the FAB report, there are undoubtedly going to be important implications for related fields and departments. Despite the call for depth, breadth, and richness in the FAB report, the committees feel that the structure as stated is primarily focused on building faculty in two units, Statistics and Electrical Engineering and Computer Science. Thus, there was concern that the plan lacked evidence of breadth and richness. While individual faculty members may have an affinity for data science in their research, the committees believe it is important that Berkeley maintain its strength in

disciplinary areas (e.g. in the social and environmental sciences). The idea that faculty across the whole campus would be invited to throw in their lot with a new school could have an incredibly disruptive and potentially damaging effect. Further, it is difficult to discern how the new structure will alter the distribution of majors and graduate students in existing disciplinary departments or units. Splitting social science disciplines between 'traditional' departments and the new unit (i.e. having a Political Science Department in one place on campus and a cluster or department of people doing data-driven political science research and teaching in the new unit) splits the discipline artificially based on methods used, not the intellectual focus of the field.

Principle 7: Striking the balance between pursuing new initiatives (in general, but for now in the particular case of data science) and maintaining strength in the disciplinary departments, majors and minors will be a critical aspect of the formulation of the detailed implementation of a structure to pursue data science on campus. It dictates that attention must be paid to the external effects of such actions on other departments, not just in terms of number of majors or resources, but also in terms of its intellectual and academic niche on campus.

Educational Component:

It was noted by several committees that the report did not contain a significant description of the undergraduate program in Data Sciences under the proposed new structure. However, it was noted by some who have been long acquainted with the Data Sciences initiative that the undergraduate educational vision was laid out in a prior report (DSERAT report, 2015), which was not included with the supplied materials and not explicitly part of the official charge. Hence the committees considered the information provided in the 4 documents supplied.

- a) Undergraduate and Graduate Curricula:** While we recognize that the curricula were not the focus of the FAB report, the committees believes that there are curricular structures that could be created to emphasize the “Breadth” aspects of Data Science without formally establishing it within the new unit. One approach to create this bridge would be for the new Data Science major to be defined by a “core plus application” structure: majors would develop a depth of expertise in data science methodologies and approaches, but would also choose a pre-defined cluster in an application field (with courses that are disciplinary – not necessarily data-centered); each application field could even define a distinct major within data science (i.e., majors in “Data Science in Sociology”, etc.). A possible model lies in the pre-defined double majors that Stanford has developed in Computer Science in conjunction with the Humanities. Nevertheless, important curricular questions remain, such as: What would be the lower division requirements for majors of Data Science? How would they compare with Engineering and L&S requirements?

Principle 8: Although the undergraduate and graduate curricula are not well described in the FAB, the senate is aware that if the initiative goes forward, COCI, Graduate Council and Undergraduate Council will have the opportunity to weigh in on details as they relate to the educational components. Therefore, the Academic Senate expects to play a significant role in development of these new curricula.

Resources:

It is important for the campus to be able to make strategic decisions to invest strongly in compelling areas that further our academic mission. In discussions on resources all committees agreed that the unit will need to be resourced appropriately with FTE, TAS, and fundraising support. Concerns were voiced about using existing resources in the current budget climate to focus on this effort, but also acknowledged that this presents a significant fundraising opportunity.

Principle 9: The FAB calls for \$2.5 million to launch the planning stage. Given the implied promise that DSI can leverage large external donations, it seems appropriate that the campus would provide, something on the order of half of the funds with the other half to come from private philanthropy. This would be an initial (very small) test of external interest.

- a) **Resource planning.** Any actual proposal of a new administrative unit requires a much more detailed study of the cost of its creation and operation - with particular attention paid to new administrators. It is possible that the marginal costs would be low but this cannot be assessed from the FAB report. We should be very cautious of creating another dean in our already crowded Council of Deans. We should really be thinking of ways to reduce the number of deans.
- b) **Funding for the new unit.** Many raised questions regarding how the new unit would be funded, particularly the balance between new funds and existing campus funds, and how this mix would be determined. The committees expressed that it is impossible to decouple the financial considerations from the structural ones. There was concern regarding how the DSPI funding would potentially impact funding for existing units and other initiatives on campus.

Principle 10: If the campus proceeds with a significant effort in this space, the new unit needs to be given the resources to succeed, as there is a record of efforts that did not succeed because of insufficient resources being dedicated to them. Given that, the committees are wary about the campus incurring costs without valid financial projections or plans to re-allocate resources should the unit not be reasonably self-supporting. However, given the optimism about the level of philanthropy expressed in the FAB report, the suggestion from some committees that the majority of resources for this new unit should come from

external sources to allow the campus to address its current short run financial challenges, could be implemented.

- c) **Instructional Resources.** Assuming that the new unit will be in charge of a new Data Science major, minor and perhaps DE, with the expected interest among our students, it is important that TAS monies be directly allocated for instructional support in proportion to the student credit hours. It was the feeling of the committees that the education program is already happening and will expand independent of what the campus does in response to the FAB report. In this time of diminishing TAS budgets, their distribution becomes increasingly important.

Principle 11: TAS funds should be directly allocated to the new unit for instructional support in proportion to the student credit hours. The implications of different institutional structures on TAS distribution should be explicitly defined and evaluated as early as possible in the process, including consideration by UGC and GC. In fact, we suggest that the campus take advantage of this opportunity to rationalize its TAS allocation policies for all units.

Principle 12: As part of any detailed planning exercise, there should be a well-articulated plan that shows how enrollment in the new major will be gated to match the available teaching resources. While it is very egalitarian to want to enroll all-comers, it is not realistic for majors in demand. It threatens the concept of a comprehensive university and would likely lead us towards a model of an institute of technology.

- d) **FTE allocation.** We expect FTE allocations to proceed as they have historically, and that they will be informed by decisions regarding endowed FTE. To provide maximum visibility to the new Data Science unit, the campus may decide to advertise that a large number of FTE allocations for the new unit are expected to be sought at the outset for the purpose of making a strong statement and displaying campus commitment. Leaders of the unit would then be expected to raise a certain number of fully funded FTEs per state FTE -- a 2 to 1 ratio was discussed as one model. Note that the funding levels for these fully endowed FTE should account for the generally high salaries in the DS fields. The unit would also need an ongoing mechanism for making its FTE requests -- something that could potentially be arranged through deans, directors or directly with the Provost.

- e) **Endowed FTE:** The proposal is, to some extent, predicated on acquiring a large number of endowed FTE. Questions arise such as: What should be the basic parameters of a formal (not an *ad hoc*) policy for handling endowed FTE requests? How much endowment is required for faculty in Data Science? If a unit finds a donor for an endowed FTE, how should that change its future FTE allocation from campus so that units have incentive to find donors, but the campus also benefits? Should the policy be uniform

across campus? One proposal is that the denominator in SCH/FTE and majors/FTE calculations (used as a rough guide when allocating new state-funded FTE) should be $X+bY$, where X is the number of state-funded FTE, Y is the number of non-state-funded (e.g., endowed) FTE, and b is a weighting factor between 0 and 1; perhaps 0.5.

Senate:

Ongoing Senate engagement. Potential external impacts of any new institutional structure pose a challenge for campus that merits deliberation. It is important that impacts be carefully considered and we recommend active and dynamic engagement with the Academic Senate during the development process.

Principle 13: The committees propose the following specific actions:

- *Formation of a senate taskforce that will interact actively with the administration as issues emerge to provide rapid-response feedback.*
- *Senate committees must be consulted as plans emerge in order to provide due consideration of the external impact of actions. For this to be successful, senate committees must also do their part to be responsive to requests for comments and analysis.*