



August 14, 2023

Administration's Response to April 17, 2023 Report on the Possible Formation of a Department of Data Science

The administration thanks the members of the Proposal Committee for a Department of Data Science ("Committee") for their careful work exploring whether UC Berkeley should establish a Department of Data Science within the College of Computing, Data Science, and Society (CDSS) and, if the campus chose not to, in what other ways might Berkeley be a leader in the area of data science and related fields. The administration appreciates the Committee's cogent summary of its work contained in its report of April 17, 2023. The administration also thanks the Committee's members for the helpful debriefing meeting held on June 29, 2023.

This document represents the administration's thinking about possible next steps.

Based on the Committee's report and the debriefing meeting, the administration finds that there are five central objectives that are worth pursuing going forward:

1. Enabling Berkeley to keep at the forefront of research that lies on boundaries between fields that is best advanced by new methods in computing and data science, ideally in ways that furthers innovation in those methods.
2. Facilitating the recruitment of scholars, especially at the junior level, working at these boundaries and ensuring their success.
3. Ensuring that "and Society" is incorporated into CDSS in a robust and meaningful way.
4. Advancing undergraduate education, particularly, but not exclusively, in the data science major, so that undergraduates not only achieve appropriate mastery of skills in computational and statistical methods, but also have meaningful experience with their application to societal, scientific, and other substantive problems and questions.
5. Establishing paths in graduate education that will produce the next generation of scholars in these areas.

Working toward those objectives first requires understanding the extent to which Berkeley's existing structures and practices may be failing to advance our achieving those objectives adequately or, perhaps, may even be inhibiting our doing so.

In both its report and in the debriefing meeting, the Committee expressed a view that Berkeley's structures and practices were, in particular, inhibiting its success with respect to objectives #1 and #2.

On page 6 of the report, the Committee writes that it “reviewed a number of examples of exceptional scholars who didn’t fit neatly into the intellectual and research priorities of existing UCB departments.” While not explicitly stated, the implication is that Berkeley failed to hire those scholars, to its detriment and at odds with the first two objectives. In the debriefing meeting, Committee members offered, as an explanation to why Berkeley’s failed to hire those scholars, that those scholars’ work was not sufficiently central to the disciplines that define our existing disciplined-based departments for those departments to be willing to allocate scarce FTEs to their hiring. A perspective underscored in the Committee’s report: “scholars of this type are not prioritized in the current UC Berkeley ecosystem of faculty searches” (page 6).

Accepting both that there is a problem and the above diagnosis of it, the question becomes how to fix it. In its report, the Committee proposes that the campus do so by establishing “small, intellectually focused, faculty-led programs in emerging fields of study that intersect with data science” (page 8). In particular, the Committee proposes a process similar to that used 15 or so years ago when the new initiative centers (NICs) were established, leading to some number of new units, which the Committee envisions as augmented graduate groups (AGGs).

While that proposal is intriguing, the administration nevertheless has serious concerns about it. First, the campus’s experience with NICs has been mixed: while some enjoyed considerable success (very notably so in the case of the Center for Computational Biology), others did not do as well. Of course, one might argue that Berkeley should be willing to experiment in such a fashion, accepting that not every experiment will be successful. Such an approach is, the administration notes, easier to entertain when the campus’s finances are rosy, which does not describe the current situation¹. In addition, independent of possible success, it is to be remembered that every unit entails certain overhead (fixed) costs; so, *e.g.*, the total cost for five units that each have five faculty is considerably more than one department of 25 faculty. A further financial concern is that, at this time, for various reasons the campus needs to be especially attentive to undergraduate education; adding graduate groups, augmented or not, does not further that priority *per se*.

In addition to financial concerns, critical though they are, AGGs may not, in themselves, address the concern that the “ecosystem of faculty searches” creates issues for hiring scholars working at or across disciplinary boundaries: some *sui generis* exceptions notwithstanding, it has not been campus practice to allow AGGs to hold 100% of a Senate faculty member’s FTE. With rare exception, all members of AGGs hold joint appointments in one or more of the campus’s departments or schools; moreover, those departments or schools are the home units for the faculty members. While a department or school might be more receptive to hiring the sort of scholars at issue if doing so “costs” only half an FTE rather than a whole FTE, recent experience has seen units balk at even such “cut-price” appointments; many recent searches joint across departments and schools, including those in the computing and data science areas, have floundered for many of the reasons discussed in the Committee’s report.

¹ It may be worth noting that the NIC process was put in motion prior to the Great Recession, at a time when the campus’s financial picture looked quite good.

While appreciative of the Committee's creativity in proposing an open call for "NIC"-like entities, as well as the thought behind the eight draft proposals for such entities, the administration respectfully believes that different approaches should be explored. Below are two alternative structures the administration thinks worthy of consideration, a department or a cluster.

A department. One alternative approach is to return to the idea of a new department within CDSS. The administration understands that the Committee was cool to the idea of a single new department: the Committee's thinking captured by "What is abundantly clear is that the model of 'one Data Science department' as the locus of all data[-]science activity would be counterproductive and undermine the remarkable progress made across the campus to date" (page 4). After discussions with the Committee, as well as other conversations, the administration's understands key factors that lead to the Committee's coolness to the idea were:

- The challenges that would ensue in distinguishing a department with the name "Data Science" from other units on campus, principally the Department of Statistics, as well as, to a degree, the School of Information ("I School"). It was noted that a number of statistics departments around the country have recently changed their names to include, in some way, data science.
- Having a department with that name would create confusion vis-à-vis management and "ownership" of the data science major.

There may well be other reasons for the Committee's coolness to the idea, including simply a preference for NIC-like entities. At the same, other than the sentence quoted above, the Committee's report does not devote space to making the case against having a single department, so it is not clear what those other reasons, if any, might be.

The issues of name and the relationship to the undergraduate major seem, in themselves, far from being insurmountable matters. Notably any name that did not include "data science" would eliminate issues vis-à-vis the data science major. Moreover, other names could also both make clear the distinction between this new department and existing units, as well as signal its "big-tent" nature, making clear that it can serve as a home for scholars working at disciplinary boundaries.

One possible name is the Department of Computing and Quantitative Methods for Science & Society ("CQMSS"). A variant on that name would replace "Computing" with "Computational." The "for" in the name indicates a faculty interested in the applications to which their development of methods based on computing and quantitative methods would be put. The "Science & Society" component makes clear that areas of application are important *per se*.

Admittedly, in many disciplines there are scholars who develop quantitative methods for that discipline (econometricians within economics would be one such example). At the same time, those scholars typically (i) come out of the discipline itself (*e.g.*, with rare exceptions, econometricians hold PhDs in economics) and (ii) they tend to focus within the discipline, with limited opportunity or encouragement to reach across disciplinary boundaries to explore questions of broader interest. So, while some may question the need to include "QM" in the department's name, there could still be strong rationales for doing so, especially for attracting and housing scholars working across

boundaries. In addition, the “computing” or “computational” part of the name signals that members of this department will be developing and applying tools from artificial intelligence and other innovations in computer science that are not, now, mainstream methods in most disciplines, even traditionally quantitative ones.

Although there would be many details to which attention would need to be paid, such a department seems, on its face, able to be sufficiently ecumenical that it could encompass a good fraction of the eight or so areas discussed in Appendices A–H of the Committee’s report. Having scholars in those areas in one department would appear to offer the added benefit of facilitating cross fertilization among those areas, something that would be more difficult to achieve if they were wholly separate entities. While tensions can arise in departments in which the faculty engage in different areas of research, there are many examples of departments and units on this campus that have made that work, some by adopting divisional structures within the department (*e.g.*, the Haas School’s nine subject-matter groups, ESPM’s three divisions, or MCB’s five divisions). Moreover, divisional structures need not remain stagnant. Indeed, Haas’s group structure has changed over time and MCB is in the process of adding a sixth division, in response to new developments, some of them interdisciplinary, and for the purpose of generating new opportunities for research and student preparation.

While such a department would not be the home of the data science major, it presumably would contribute to it in important ways: first, presumably a number of its faculty could teach core courses of the major; second, these faculty would be well placed to develop important connector courses; and, third, they could contribute to a number of the major’s domain emphases.

In terms of graduate education, a CQMSS department could be an umbrella for a number of doctoral programs, such as the PhD in Computational Biology, or it could develop a broader program in which students might pursue emphases.

As noted, details remain, including who the faculty would be. One question is whether the Center for Computational Biology (CCB) would be a part of the new department. Its being in CQMSS would provide CCB the benefit of a departmental home and could help address issues it currently confronts. It would also provide CQMSS an immediate mass of faculty. A possible downside—or at least a challenge—is avoiding computational biology dominating CQMSS in ways that kept it from becoming an ecumenical home for scholars working on the boundaries of computing, data science, and other disciplines (*e.g.*, the social sciences, the humanities, and other natural sciences). Another source of faculty in the near term for CQMSS would be faculty moving a portion of their FTE from their current units to CQMSS. Mechanisms to govern such movement, as well as appropriate campus-level review to ensure existing units can remain vibrant despite the loss of some (fractional) FTE will be necessary. Additionally, after appropriate campus-level review, it is possible that a small number of new state-funded faculty positions could be allocated to the new department, positions that could be used either for 100% or joint appointments within it. That said, budget realities suggest that it would be challenging to allocate many state-funded faculty FTE to any new unit. Finally, it is possible that philanthropy could provide support for faculty positions; although caution is necessary insofar as both the campus and CDSS arguably have much higher philanthropic priorities at this time.

While financial resources should not be the primary driver of academic decisions, we must consider finances in our decisions. It is the sense of the administration that it would be substantially more affordable to support a single new department with N divisions than it would be to support N new AGGs.

A “cluster.” Another alternative is some variant on a cluster-hire initiative. As the committee knows, the campus has had a series of cluster-hire initiatives (*e.g.*, Climate Equity & Environmental Justice, Latinxs & Democracy, etc.). One goal of such initiatives is to provide incentives to existing units to recruit scholars in areas in which they might not have otherwise recruited. Additionally, by recruiting the cluster as a cohort, with some thought to facilitating interactions among those hired in the cluster, a cluster can be more than the sum of its parts, even if its members are spread across different units.

Advantages to cluster or cluster-like initiatives are:

- They are easier to establish than formal units such as AGGs and departments;
- They avoid the costs associated with establishing and managing formal units;
- They are low cost in terms of being experiential insofar as if they don’t work, the faculty are simply “absorbed” into existing units; and
- They augment rather than draw resources from existing units.

The principal disadvantages are:

- The absence of a formal structure, such as an AGG or department, can make it challenging to have them be more than the sum of their parts;
- Because those hired are in existing departments, it can be challenging to have those hired contribute to a common good (*e.g.*, augmenting the data science major);
- The *direct* benefits for undergraduate education have, so far, not been large; and
- Faculty hiring often requires coordination among departments, which has, in many instances, proved challenging to affect and maintain.

One can conceive of a number of variants within the cluster or cluster-like paradigm:

1. Similar to what the Committee recommended, but for clusters rather than AGGs, a call for proposals could be issued, in which the campus committed to some number of clusters pursuing different flavors of computing/data science plus X, each cluster having some number of faculty FTE attached. Two, somewhat related, challenges with this variant are that without added incentives, the clusters don’t *per se* overcome the “ecosystem of faculty searches” problem; on the other hand, the typical incentive—having the FTE float against units’ guaranteed minima of state-funded Senate FTE—raises the cost of the clusters, especially if

the floats are long enough to provide strong incentives, which in turn would limit the number of state-funded FTE that could be allocated to these clusters.

2. A central non-departmental entity (“center” or “institute”) is formed and it is allocated some number of faculty FTE, for which it, in turn, conducts searches in areas of computing/data science plus X. Candidates identified in these searches are then offered to departments or schools for appointments. These positions would not count against the appointing department(s) or school(s) guaranteed minimum of state-funded Senate FTE, although when an individual appointed to such a position separates, the FTE is not kept by the appointing department(s) or school(s), but goes back to the central entity². A downside to—and very significant concern with—this approach is that searches conducted in this manner have often floundered when it came to housing the identified candidate. To an extent, this downside can be mitigated via the appropriate construction of search committees.
3. A variant to #2 is to have departments and schools, as individual units or in consortia, propose to the central entity searches that would make use of the FTE allocated to the entity. The entity, perhaps with further central-campus review, approves or denies these requests. The FTE accounting is the same as in variant #2, as is “ownership” of FTE upon separation. It is conceivable that the entity could also grant FTE “*ex post*” if, say, a unit’s disciplinary search uncovers a candidate who would be a welcome addition to that unit, but, because their work is in computing/data science plus X, they might be more attractive to the unit if their FTE did not count against the unit’s guaranteed minimum³.

Because the candidates hired in a cluster or cluster-like model would not necessarily be in units within CDSS, how they might contribute to the data science major is less clear than if they were in a department within CDSS. Hence, in terms of objective #4 above, a cluster or cluster-like model might advance undergraduate education less in data science *per se*, although being embedded in departments could mean faculty hired under such a model are well positioned to provide students a better understanding of the application of computing and data science methods to societal, scientific, and other substantive problems and questions than were they in a separate CQMSS department.

Budgetary considerations. As the Committee recognizes, the campus faces significant budgetary constraints. Notably, from a financial perspective, maintaining the roughly 1500 state-funded Senate FTE we have has, at times, proved challenging budgetarily. Although the sum of the guaranteed minima for state-funded Senate FTE is 1425, providing a 75 FTE “cushion,” the cushion has been largely used (*i.e.*, there are incumbents in those “cushion” slots). Moreover, as important as scholarship in the areas under discussion is, there are many other important initiatives that the campus also wishes to advance. Consequently, allocating ten *new* state-funded Senate FTE to this

² To avoid moral hazard problems vis-à-vis tenure decisions, there may have to be some special rules to govern separations resulting from tenure denials.

³ It should be noted that such *ex post* slot allocations could likely prove problematic in practice for a number of reasons, including issues of search equity, timeliness of action, and concerns about gaming by departments and schools.

purpose, as the Committee recommends, is a large ask. Moreover, even the 15 philanthropically funded FTE envisioned by the Committee represents an additional five state-funded FTE if one imagines the 15 philanthropic FTE would be under the Chancellor's FTE Chair model, in which each new FTE is two-thirds funded philanthropically and a third is "funded" as a state FTE. Because 15 state-funded FTE (10 + 5) is just 1% of the campus's state-funded FTE, one might, from that vantage point, reasonably argue that what the Committee seeks to allocate is not a significant investment vis-à-vis the potential return. The administration's perspective is that we must, instead, focus on the margin, in which 15 FTE is a fifth of our already small (and non-empty) cushion.

There are, of course, other budgetary pressures that will make it challenging to accommodate significant expansions in staffing and graduate student support.

Consequently, in any reformulation of the Committee's proposal or any other plan that is put forward, it will likely be necessary, at least over the near term, to ensure that what's proposed fits within the campus's financial reality. Even if one is optimistic about philanthropy in this area, it needs to be remembered that philanthropy may not be without opportunity cost: we must be cautious about gifts that come at the expense of obtaining funding for higher or more-pressing campus needs.

While the administration has concerns about the details of the Committee's proposed means of addressing objectives #1-#5, the administration agrees with the Committee that it is important that Berkeley address them, at least to the best of its ability. The administration welcomes the Committee's thoughts, as well as those of others, as the campus continues to work on this critical matter.