

To: Scott Strobel, Provost
Michael Crair, Vice Provost for Research

From: Data-Intensive Social Science Task Force
Joseph Altonji, Economics
Judy Chevalier, School of Management
Alan Gerber, FAS Dean of Social Sciences; Political Science
Susan Gibbons, Provost's Office (chair)
John Lafferty, Statistics & Data Science
Tim Pavlis, Provost's Office
Barbara Rockenbach, Library
Jasjeet Sekhon, Statistics & Data Science; Political Science
Megan Smith, School of Medicine; School of Public Health
David Wilkinson, Tobin Center for Economic Policy

RE: Task Force Recommendations

Date: February 18, 2021

The University Committee for Data-Intensive Social Science (DISSC) [report](#) recommends establishing a university center ("Center") to support advanced computational and data-intensive methods for social science research. This Center has three main objectives. It will seek to provide the physical and human infrastructure to facilitate research projects based on acquisition, computing, and storage of sensitive, complex, and restricted-use data; promote data-intensive social science research through sharing information about data resources and methodological innovations, providing specialized training and consulting on computational and statistical methods and raising awareness of Yale's infrastructure to support data-intensive research; and seek to build community and support research connections around data-intensive methods across the university's departments and schools.

In November 2020, Provost Strobel charged our small task force to recommend concrete steps for the creation of the Center. The Data-Intensive Social Science Task Force met between November 2020 and February 2021. The task force's primary goal was to outline the initial actions that will result in a successful data-intensive social science center (name to be determined), leaving many of the key decisions, such as staffing and programming, to be determined by the Center's inaugural leadership.

[Recommended initial actions](#)

Appoint a faculty director for the Center: The faculty director will provide the intellectual and academic vision for the Center. A critical early task for the faculty director will be to oversee the appointment of the Center's executive director (title to be determined). The Tobin Center and Tsai CITY provide two successful examples of the faculty director/executive director partnership.

The faculty director is critical in providing the Center with the "institutional navigation" necessary for success at Yale. In addition, the faculty director will chair the Center's university-wide advisory committee. While the faculty director will inevitably come from a particular academic department and school of Yale, the ideal candidate for this position embraces an understanding that the Center is designed to support data-intensive social sciences across all academic departments and schools of Yale.

The Center will be well-positioned for fund-raising and grants, and consequently, a director with experience in those areas will be ideal.¹

The faculty director should have an appointment term consistent with other faculty directors of centers at Yale, with a full review in the penultimate year.

Appoint an executive director for the Center: A critical early task for the faculty director will be to oversee the appointment of the Center's executive director. The executive director oversees the daily operations of the Center. It is the job of the executive director to design and build the Center's enterprise with a strong customer service mandate. The ideal candidate brings first-hand experience with data-intensive social science research, coupled with experience in organizational design and product management (e.g., user-focused design, build, delivery, marketing and advocacy of a service). The executive director should also be comfortable with fund-raising and grants.

Appoint a university-wide advisory committee: The DISSC report recommended the creation of a university-wide committee to share information about data-intensive social science and provide advice to university administrators and key units, including YCRC, OSP, ITS, and library, that have substantial roles in supporting data-intensive social science research. The task force recommends that this committee should be established and should also serve in an advisory capacity for the Center. The advisory committee will provide feedback on the alignment of the Center's service programming with the needs of their communities, as well as assist with "trend spotting" to ensure that the Center can be proactive to the rapid changes in data-intensive social science research. Such a committee would help in developing key performance measures in order to build a culture of organizational assessment from the start.

The advisory committee should hold as a core priority the provision of assistance and advice to the directors in executing the central 'customer service' mission of the Center. In connection, the Center should seek to continuously track and understand the nature and breadth of researcher demand for services. It should seek to use modern user experience approaches to do so. Leadership, with counsel of the advisory board, should set priorities for action without overextending and derive both milestone goals and performance metrics based on those priorities.

The members of the committee should inclusively represent the broad data-intensive social science research interests of Yale and include representation of key service providers, such as ITS, YCRC and the Library.

The Center's directors may wish to work initially with a more *ad hoc* group of faculty who are deeply engaged in data-intensive social science research to gain a better understanding of the environment at Yale. Our observation is that researchers deeply engaged in data-intensive social science are within 10-15 years from their PhDs, while university-wide committee appointments can often skew towards more senior faculty. Similarly, the directors may wish to create a smaller, on-going executive committee that can provide more frequent and in-depth consultation than the university-wide advisory committee.

¹ The ideal candidate would have both technical expertise and be able to reach out to faculty members who are at early stages of the learning curve; this may be a difficult combination of skills to find. Stanford has used co-directors to draw on different skills and networks (example is [Welcome | Stanford HAI](#)). This center is co-directed by a philosophy professor and a computer scientist.

Identify and assign a chief data counsel: The DISSC report recommends establishing a chief data counsel, who is a data-use agreement (DUA) specialist with experience relevant to data-intensive social science, to maintain critical external relationships with data partners, negotiate DUAs, oversee DUA counsel, and support University-wide efforts to minimize risk and increase efficiency. The inquiries of the task force confirm the urgent need for this position which is felt widely across social science and science departments in nearly all schools of Yale.

The successful execution of a data use agreement can often involve the work of staff from Office of Sponsored Programs (OSP), ITS, Procurement, Yale Center for Research Computing (YCRC) and Office of General Counsel (OGC). Faculty express a high level of frustration as to the opacity of the process and the time and effort it consumes. In fall 2020, Alan Gerber conducted an informal survey of faculty who regularly use data and found that the frustrations had not abated since a similar inquiry was done by the DISSC committee in 2018-19. ITS is working closely with OSP and Procurement to create a DUA portal, which would track the movement of a DUA through the review and approval process. While the portal will be very useful, and the task force supports further consideration of this approach, the portal will put an interface on top of a set of processes that are not yet working well for researchers. Yale does not have staff in the aforementioned offices dedicated to DUAs. Consequently, there is a lack of consistency of practice and the lost opportunity to build up experience and knowledge within a set number of individuals.

When seeking to secure a data use agreement from a powerful company (e.g., Google, Uber, Facebook) or from a national or international governmental agency, it is critical that Yale's researchers are represented by someone ready to operate at a similarly high level of legal sophistication. Currently, many DUAs are being handled by lower-level staff in Procurement and OSP, who are not specialists in DUAs and may not have the appropriate authority to negotiate. When a DUA is moved to OGC, it is competing for the finite attention of OGC attorneys. It is important to recognize that when it comes to DUAs, Yale is often the supplicant that must adjust to the requirements of the data owners. Junior administrators of university data policy do not often have the wherewithal or authority to engage in negotiations, to make necessary policy adjustments, and do so while protecting Yale's interests.

Between FY19 and FY20, the number of DUAs that flowed through OSP more than doubled, from 133 to 271, with another several dozen passing through Procurement. We expect such rapid growth in DUAs to continue. The task force recommends that OGC creates a new position, dedicated to the review, negotiation, and approval of social science DUAs. The chief data counsel would work closely with OSP, Procurement, ITS, YCRC and colleagues in OGC to shepherd complex social science DUAs through the review and approval process. In more complex cases, the chief data counsel would also be the consistent point of contact for Yale's researchers, who often need to remain involved in the DUA negotiations in order to answer questions about use and/or to leverage their existing relationships with the data owners. The chief data counsel should also assess the effectiveness and efficiency of the current institutional and procedural DUA infrastructure and make recommendations for structural improvements to process consistent with protection of Yale's interests.

The appointment of a chief data counsel should not wait until the formation of the Center. Such a position will have an immediate, positive impact not just on data-intensive social science research, but data-intensive research across all disciplines. Once the Center is established, it would be ideal for the chief data counsel to be at least partially seconded to the Center so that the individual can more easily

become a part of the data-intensive social science community. Since the dissatisfaction with the current DUA system is felt just as acutely by the various departments and schools of Yale, it is reasonable to assume that the cost of this position could be shared by the Center (on behalf of the social science schools and departments), and OGC and potentially the School of Medicine for those DUAs that sit at the intersection of social sciences and medicine.

Identify initial administrative/financial support: The Center will require some level of initial administrative support, which will grow over time. In addition, the Center will need a financial home and access to a business operations team. There is likely the opportunity to cost share with other departments or centers for this support in the early stages. Such partnerships may be driven by the decision of where to place the Center, organizationally, within Yale.

Other Center staffing:

Data Engineering/Research Consultant: The DISSC report also recommends a data engineer to coordinate the technical elements of data acquisitions and data security, as well as a research consultant who can assist with data discovery and use. The YCRC provides the model of a research support analyst who is dedicated to a particular unit or discipline, in this case the social sciences. Such joint appointments have been successful between the YCRC and the Wright Lab and the Yale Center for Genome Analysis. It was unclear whether the tasks of the engineer and consultant can be successfully combined into a single person, and the task force feels that this is a question that should undergo further exploration with Kiran Keshav (YCRC) and the Center's faculty and executive directors. Moreover, in developing the portfolio of services, the Center's directors may see the need for a different type of role, such as a data editor. Finally, these positions should bring complementary skills and experience to that of the Center's executive director. For these reasons, we recommend that two FTEs be set aside for the future use of the Center, starting in the second half of FY22.

Other ideas for consideration by the Center directors:

External inputs: Many peer institutions have created similar data centers or are in the development process. Forming an external advisory committee made up of representatives from peer institutions establishes a formal and early conduit through which we can learn more about best practices and lessons learned.

In place of an external advisory committee, or in addition to, the DISSC report recommended that Yale's Office of Institutional Research and Strategic Analysis (OIR/SA) produce a periodic memo on "Innovations and Lessons," perhaps annually, on major developments supporting data-intensive social science research and teaching at industry-leading organizations and peer universities. The memo would describe important infrastructure investments, important changes in data policies and data availability, key programs being started, and programs being discontinued. This memo should be sent to university deans, center directors, department chairs, and shared with the faculty. Based on this research, OIR/SA might propose one or more faculty site visits each year to places that seem especially innovative or cost effective.

An alumni advisory committee: Yale has many alumni who are leaders or investors in the digital economy and are engaged in data-intensive social sciences, and/or are advocates for its growing importance. The network that such a committee would create could foster data acquisition, smart purchasing, trend spotting, and open doors. It could provide a distinctive new outlet and opportunity for alumni to

support a Yale activity close to their passions and expertise. Many units of Yale have an alumni development council through which many key alumni donations are solicited. With the upcoming capital campaign, this committee could be a force multiplier for fundraising.

Technical consulting: Depending on the background and expertise of the directors as well as the priorities they set, they may value targeted and technical consultation services. This, for instance, may include UX (user experience) experts to define researcher needs with greater depth and breadth; digital strategy consultation on cadence and prioritization of tasks; independent, third-party services on smart procurement; technical personnel scoping and search; assistance in development of peer a rich analysis or other landscape research; targeted technology, product management or systems trainings.

Stat Lab: The StatLab is a unit within the Marx Science & Social Sciences Library which provides support, training and consultation for data analysis and statistics. The DISSC report cites advantages to relocating Statlab to the proposed Center. The location of the Statlab, the service model, and management and governance of the Statlab should be an early conversation between the Center's directors and the University Librarian. Relocation of these consulting services to the Center would make it easier for social science faculty to ensure coordination of services with faculty and student needs, provide greater oversight and quality control, and increase convenience of service delivery due to closer proximity to the center of gravity of Yale's social science faculty and students.

Space:

The task force was very pleased to learn that the Provost decided to assign 85 Trumbull Street, the building that was recently relocated from 87 Trumbull, as the Center's interim location. In the summer of 2023, when the Economics department moves to its new building, the Center can move to 37 Hillhouse as its long-term home.

This timing allows the Center directors the opportunity to map the Center activities against the affordances of the space within 37 Hillhouse and make any renovation suggestions in advance of the move. The building has 26 offices and 6.5k assignable square feet, which will exceed the Center's initial space needs.