Academic Data Science Alliance

Organizational Structures Panel Report

2018 Data Science Leadership Summit

Melissa Cragin and Tyler Kloefkorn

Executive Summary

Data science is fast emerging as a new discipline. Data science initiatives across universities and colleges in the US and beyond are emerging at a rapid rate. That data science is as much a practice as it is a discipline, raises challenging organizational questions - for example, whether and how data science should be its own major, department, or division at a university. If data science is siloed into one academic unit, how does one effectively integrate data science skills and best practices into curricula in other departments, particularly those necessary for students in other disciplines to reach the expectations of 21st century employers? This panel was an outgrowth of discussions held during the inaugural Data Science Leadership Summit held in March 2018 at Columbia University, with the goal of furthering understanding of the approaches to data science developing at US institutions. For this Summit, the Program Committee, led by Srinivas Aluru, chose “Data science organizational structures” as one of the current challenge areas for academic departments and cross-departmental initiatives. The session was organized to meet the objectives of the Summit and learn from those leading exemplar programs about success, challenges, and lessons learned. The goal of the panel was to move beyond the structured information already gathered on entities and their institutions (from the Summit report) to questions and decisions encountered along the way. How do we make data science entities work for our institutions? How do we make them sustainable? What approaches seem to be working well? What are the challenges?

Key topics covered include:

- Models of campus engagement
- On-going challenges for the DS entity at the Institutional level
- Key tactical decision points - perspectives from “looking back”
- What’s important as the DS community moves forward?

Significant take-aways:

- A number of institutions have achieved considerable success in establishing a “virtuous cycle” in which the “producers” of data science methodology are closely partnered with the “consumers” of that methodology, with advances in methodology driving advances in discovery, which in turn drive further advances in methodology. While the institutions belonging to the Moore-Sloan Data Science Environments project (Berkeley, NYU, and the University of Washington) had a head start in this regard, many other institutions, such as Columbia University (started in 2012 with NYC funds) have met with success.
- Many institutions are experiencing tremendous demand from all corners of the campus for data science education of all forms, ranging from consulting and tools-oriented short courses to degree programs at the bachelors, masters, and doctoral level. Education can be a good way to create connections, collaborations, and goodwill.
- Campus and local community engagement - i.e. active stimulation events or interactions - is a significant aspect of most of the institutional models discussed on this panel. While timing and processes vary, there are widespread efforts to bring together faculty, students, and projects to leverage expertise and resources. Differences in the ways that research and scholarship are funded across domains require different approaches to bringing together faculty “under one roof.”

- Variation in institutional and domain-based incentives, as well as compensation (e.g. grant-funded or not), will impact the level of interest and engagement across departments and faculty.
- Only a small number of institutions are organized to generate revenue.

Editorial Note: Some sections have been modified by panelists to update perspective and provide additional context. We are grateful for these additions. (March 12, 2020)

I. Opening session

Context: Ed Lazowska opened the session with a short presentation to frame the session, starting with some background on the Moore-Sloan Data Science Environments (MSDSEs). These "environments" were funded about six years ago, and are hosted at the University of Washington eScience Institute, the Berkeley Institute for Data Science, and NYU's Center for Data Science. MSDSEs connect data science methodologies to scientific theme areas and they intend to transform discovery in all fields. It was noted that all three organizations have expertise within their environments covering both data science methodologies and scientific theme areas. The goal is to create a "virtuous cycle" where advances in data science methodologies lead to scientific discoveries and scientific discoveries lead to advances in data science methodologies.

MSDSEs are not driven by methodology disciplines (e.g., computer scientists and statisticians); rather, these entities are true partnerships that are careful to avoid "ownership" of data science by any single discipline or small collection of disciplines. Recently, a whitepaper was written by the three MSDSEs, aiming "to provide a menu of possibilities for those seeking to emulate aspects of our Data Science Environments":

- The Moore-Sloan Data Science Environments, "Creating Institutional Change in Data Science": https://doi.org/10.6069/v5s4-8n41

Additional, related reports released during the past year include:

- ABT Associates, "Academic Data Science Centers in the United States" (October 2018): https://doi.org/10.6069/H63V-V078
- The National Academies, "Data Science for Undergraduates: Opportunities and Options" (2018): https://www.nap.edu/catalog/25104/data-science-for-undergraduates-opportunities-and-options

Presentation on findings from inaugural Summit: Jeannette Wing made the next presentation, in which she summarized the DS Summit that was held in March 2018 at Columbia University. This previous, inaugural meeting was intended to identify data science entities - that is, initiatives, institutes, departments and divisions - that are emerging across the higher ed landscape, with particular focus on the top research universities. The goal of the workshop was to learn what is happening with these entities, and their institutions. This Summit was inspired by the CRA/Snowbird model, in which Computer Science department heads meet biennially to talk about issues and share best practices. Also of interest is the CRA Taulbee Survey which annually tracks and reports data on CS in higher education, and how this approach might be applied to the emergent Data Science space. Dr. Wing relayed findings from the meeting, from which it was clear that one data science model will not fit all institutions (summarized in the Wing et al. 2018, Section 2.2).

This previous Summit was launched by asking, “How does data science fit into a university?” Participants described the unique circumstances of their entity and institution. Responses were numerous and complex, which resulted, notably, in changes to the meeting agenda. **The group decided to start building a taxonomy for the organization of data science efforts in higher education.** A university administrator or director looking to build a data science entity could use the taxonomy to consider variety and high-level decision points. As an initial step towards such a taxonomy, a survey was conducted during the summer of 2018, following the Summit; results were included in a workshop report (Wing et al. 2018). Attendee comments and survey responses varied but several models were identified:

- brand new unit (e.g., the Division of Computing, Data Science and Society at UC Berkeley);
- repurpose existing unit (e.g., Statistics and Data Science at Yale University);
- create a new standalone institute (e.g., the University of Washington eScience Institute and the Columbia Data Science Institute);
- create a new overlay entity that is joint between multiple academic units (e.g., the Institute for Data, Systems, and Society at MIT); and
- some combination of the above models.

It was noted that some institutions have multiple entities, with different (and hopefully complementary) missions.

II. **Panel and Discussion:** Introducing the next segment, Melissa Cragin gave a brief synopsis of the aims of the panel for this session, which was intended to move forward from the structured information we have on entities and their institutions (e.g., the Summit report), to greater understanding of on-going challenges and decision points encountered as entities grow and evolve on campus. We anticipated that this content could serve as guidance to others who are ramping up or fine-tuning data science efforts on their campus. The panelists were asked to consider several questions in preparation for the session, including:

- How do we make data science entities work for our institutions?
- Where is data science thriving and what elements are correlated with that?
- How do we make them sustainable?
- What has worked well; and, what are the challenges?

Panelists were drawn from institutions that have established or growing data science initiatives, and which have different models of program development and processes. This group included:

- **Brian Athey** (University of Michigan)
- **Magda Balazinska** (University of Washington)
- **Cathryn Carson** (University of California, Berkeley)
- **Brant Cheikes** (University of Massachusetts Amherst)
- **Julia Kempe** (New York University)
- **Saul Perlmutter** (University of California, Berkeley)

Each panelist opened with a brief description of the organizational structure of data science efforts on their campus.[1] Following these introductory statements, the panel was asked to address a set of questions that had been provided ahead of the meeting. The aim was to illuminate the ways that DS programs were evolving, including panelists’ perspective on successful and challenging aspects of program implementation. In this section, responses are

[1] These brief descriptions were not captured for this report; however, the Abt (2018) report provides detailed information on the Data Science efforts at the majority of the institutions represented here.

tallied by institution to illuminate a coherent picture of the drivers and challenges for these distinct entities.

What's Working: Panelists were asked to describe how the DS entities at their institutions are working, and related affordances.

The University of Washington eScience Institute has a collaborative structure, led by a faculty Director, a Director of Research, an Executive Director, and an Executive Committee, and staffed a team of research and data scientists who are experts in data science, cloud computing, and a variety of physical, life, and social science disciplines. There is a large number of affiliate faculty, but there are no faculty lines housed within the Institute; rather, the Provost provided a number of "half slots" to be allocated to departments hiring individuals with deep capabilities both in data science methodology and in some application field. Postdoctoral fellows and graduate students are funded by different grants received over time, and have opportunities to participate in various mentoring activities (e.g., career advice, career fair). The eScience Institute includes a partnership with the UW Libraries and is housed in former library space on the top floor of the Physics/Astronomy Tower, sending a clear signal that the effort is about partnerships, rather than being narrowly focused on and/or owned by the methodology fields. Emphasizing this, the original Executive Committee included 8 individuals from 8 different academic departments representing 4 different schools/colleges; the original 8 have been joined by 7 others representing 4 additional academic departments; the original Director was from CS, the current Director is from Astronomy.

The Center for Data Science at the University of Massachusetts Amherst is set up for "decision agility," giving leaders the ability to launch new activities quickly and serve constituents without "a ton of approvals." Naming the Center as, "Center for Data Science" gave it excellent visibility, and this results in a lot of engagement with campus constituents, and new connections.

The Center for Data Science at New York University has a robust revenue stream **and** agility through its educational structures. The Center provides many services, including seminar series, carpentry workshops, etc. There are many connections to various schools, departments, and libraries.

The University of Michigan's Institute for Data Science (MIDAS) has generated several "Challenge Centers" that are seen to be very successful, the most recent one was launched on data science and music. While the Center's funding has cost $11M, the Institute has raised $30M in new grants. The education program is reportedly going very well, and appears focused predominantly at the graduate level, and the "extremely active graduate certificate program [is] (the most successful program at the University of Michigan – it is highly interdisciplinary)." The Institute launched and supports four independent student organizations; one of these groups won KDD prizes for work on the Flint water crisis. While the Institute is hiring data scientists, it was noted that some senior data scientists are lost to industry and a few to private universities, driven by salaries, spouses, etc.

The University of California Berkeley's Institute for Data Science (BIDS) is now a unit of the new Division of Computing, Data Science and Society. BIDS provides a physical space that brings people together – it's a central place within a University library. The goal was to create a culture in which people talk to one another, and Postdoctoral fellows are very important for the culture. Typically, half of their appointments are with BIDS, and half are with their own research group. The (physical) cultural center provided a space where "people learned the language of data science." BIDS and the Statistics Department provide an academic home for large open-source software projects (e.g., Jupyter notebooks). With respect to education in the Division of Data Science and Information, there is a thriving data science undergraduate major, which is good for bringing students from different domains into the collaborative data science environment. Overall, the UC Berkeley data science entities serve a massive number of undergraduate students, and the view is that, "providing a good data science education is great motivation for the data science entities." The Foundations of Data Science class - Data 8 - has 1,300 students per semester and it is co-taught by computer science and statistics faculty members, and classes across campus are getting involved in transforming disciplinary curricula. Research across campus is reported to be thriving, also, and there are opportunities for people to be engaged in ethical and societal reflections. Berkeley is committed - and has been committed from the start - to this aspect of data science.

Particular challenges at the institutional level: This question was intended to uncover on-going "friction points," particularly as they might relate to future changes in implementation approach, or potential decision-points.

The University of Washington eScience Institute was established as a campus-wide organization with a small amount of central funding in 2008, long before the "data science gold rush." The combination of the early launch and the clear commitment to campus-wide ownership and impact has mitigated many of the institutional challenges that have been faced by data science entities that were established subsequently. Establishing meaningful career paths for data scientists is the principal institutional challenge faced by the eScience Institute.

The approach at the UMass Amherst Center for Data Science is considered "all carrot, no sticks," though it is felt that activities occasionally lack balance in terms of give and take. Everything is expected to demonstrate value to organizations on campus, including in the Center's efforts to get collaborators. Engaging "multiple college collaborators can be challenging due to communication and travel." The Center recognizes that they need to "be responsive enough to be seen as a positive force on campus," while also functioning in a cultural context with "a range of transactional relationships." Community building is an on-going challenge, and this is something the Center is actively working to improve while also avoiding "the risk of being preyed upon for service support."

One big challenge at the NYU Center for Data Science is working with faculty from diverse academic units. For example, there is a disparity in how faculty are supported - some faculty members don't need grants and some live off of grants. Working on this interdisciplinary area requires a unified balance of cultures, flexibility in decision-making and improvised solutions, etc. Open dialogue has been key. Another challenge is related to the success of the Center: other entities move to take resources. An additional challenge is distinguishing a data science entity from an existing organization, like a computer science department.

At the University of Michigan's MIDAS it is a challenge to recruit external faculty. At Michigan there are 19 schools and colleges and ten relatively new Deans, and this means that departments frequently compete with one another for faculty recruits. Decentralization is a problem, a challenge for communicating across the university infrastructure; and, "communicating in a hub and spoke model can be a challenge." An additional force against hiring external talent is competing with industry, and the Institute has lost faculty to Facebook, Amazon, Netflix, Google, etc. Some potential recruits are lost to private universities, and occasionally faculty leave due to family circumstances (spouse in another location). Providing competitive salaries tends to be an issue.

At Berkeley, organizational leadership is somewhat decentralized: “bottom up and crazy.” The organizational structure is not a tree; instead it’s a graph, and often there are negotiations across campus. Everyone has their own community and their own lines to the provost, making it more difficult to “nail down resources and authority.” Data science entities would be more successful “if they could go beyond the outdated hierarchical structures of 20th century academia.” It could be a tactical mistake to start with a master’s program and focus too heavily on revenue generation. Overall, 5,000-8,000 undergraduate students are served by these data science entities. The size is a challenge, but it is also a creative force on campus – there is no way to ignore its impact on campus.

Current status and plans for change: Panelists were asked about the current state of development and operations in terms of prior tactical choices, and success. In addition, to consider new decision-points, and potential changes in approaches to community engagement and partnership development. What are the major decisions that you faced, and that others are likely to face?

In the UW eScience Institute, there were decisions to be made on key areas of focus (e.g., methodology, or domains, or both?). There was a key decision about the Institute being a cross-cutting entity vs. some sort of department equivalent. Another key choice (at the institutional level) is whether the entity will have faculty lines or not, and whether this organization will have staff, such as data scientists or research scientists. Additional initial planning questions include, “Should the Institute be virtual and bring people together? Will it have space, and if so, where should it be located? And, how does the entity invest in postdoctoral fellows and graduate students?”

At the UMass Center for Data Science, a framing question has been, “How does an entity allocate time and resources for education, research, and service?” Every center will have an obligation to contribute to the campus mission in each of these mission areas. Early in its life, the Center for Data Science prioritized investments in the educational mission (e.g., creating a new MS-level concentration in data science as well as a complementary certificate program). As the Center matured it increased its efforts substantially to foster strong connections with industry. Industry engagement efforts have yielded new resources, but have also brought challenges associated with addressing the often bespoke needs and expectations of its industry partners. In the future, the Center will need to implement a sustainable approach to supporting the data science needs of principal investigators elsewhere on the UMass campus (i.e., outside the College of Information and Computer Sciences)..

One challenge the Center has faced is getting faculty to be engaged. The UMass Center for Data Science does not “own” any faculty members. Instead, there is a loose notion of affiliation between the Center and a subset of the computer science faculty. Faculty support of Center activities is necessarily predicated on mutual trust and a shared sense of long-term mutual benefit that must be carefully cultivated and stewarded. In contrast, the NYU Center for Data Science chose to have joint faculty lines. Initially, there weren’t enough and there were issues with service responsibilities (i.e., credit and compensation). Faculty members need perks (e.g., status), but often this isn’t enough. Overall, the center needs to create and reward stakeholders.

Michigan found there are decisions to be made about keeping faculty engaged, as well. Incentives that work for some faculty members won’t work for others. The mission that the data science entity plays - its role on campus - is the big determining factor. Faculty lines do not work well at Michigan, but certain incentives do. (Though, one approach that did not work was course buy-outs, because there wasn't enough money and it didn't apply to assistant professors.) A dream solution could be partial endowments for research appointments (i.e., have protected time for researchers). There are decisions to be made on engaging postdoctoral fellows and graduate students (and potentially undergraduate students). In these ways, through considered, context-driven decisions, "the engagement decisions can help the entity be a peacemaker of sorts on campus."

The Berkeley group has found that key decisions are set to be made about the scope of data science activities. "It's 'go big or go home' – this is the most transformative thing to hit the university in decades." Faculty know how important this is. It takes a certain level of risk tolerance, and not all university administrators have this. It took a cohort of faculty to push these initiatives through three chancellors and four provosts. Berkeley is a state school, which means that it prioritizes education. One key decision was to keep the educational responsibility (i.e., teaching) within the various departments, and not shift this to an institute, "which there is a tendency to do with activities that don't fit a department structure." The teaching activities brought departments together, and this served as an integrative force on campus. The [data science network is] "a graph – not a tree!" We don't want to be competitive.

Pathways for growth: A final question addressed by the panelists concerned approaches to broadening campus participation, especially to engage humanities, arts, and social sciences.

The eScience Institute tries to both welcome those who want to come in and reach out to those who haven't gotten engaged yet. The Institute communicates with department heads, and has gotten traction. Broad-based calls for collaborative projects are issued twice per year. Widely advertised office hours attract a broad range of faculty and students and make the physical home of the eScience Institute into a true campus center for data science exploration.

At the NYU Center for Data Science, the mission was, "to spread the data science gospel." The Center utilizes the master's students to build bridges to other departments (including social sciences and the medical school). These students are highly qualified! Sponsored research internships and capstone projects are the vehicles for this collaboration. The former are driven by department demand for student interns, and they come to the Center to do collaborative projects; the latter is driven by problems generated by departments and industry partners. Student connections can help lead to stronger connections with longer-term projects.

Part of the vision for MIDAS is, "engaging the digital humanities and some of the soft sciences - it is so important. Nurturing interdisciplinary data science teams through campus matchmaking requires a lot of investment in, and commitment to, institutional memory, energy, and will. Faculty sometimes come together around issues in fairness, accountability, and trust." A significant aspect of the work at MIDAS is conducting a lot of matchmaking - a little like the DARPA process. This work sometimes requires assessing requests, which require vetting and thoughtful responses. New efforts can start with a white paper and simple introductions. For example, PhDs who have left academia are asked to get involved. Finally, it's important to invest in PhD-level staff and to keep them engaged.

At Berkeley, calls for faculty engagement were important, and the group made a point to reach out to all disciplines. There was a multi-pronged approach: There were calls for shared data science graduate students and postdoctoral fellows with many departments; seed funding was available; there were calls for faculty affiliations; and finally, there were lots of efforts to promote integration through undergraduate teaching. It was noted, though, that, "some departments do not feel like they are part of the data science culture," making specific outreach efforts very important, "to make appointments, have lunches, and start conversations, to bring in humanities, social sciences, arts, etc., and offer leadership opportunities." A beneficial outcome of the process was "tapping into expertise in how to address organizational challenges, as they could address trust, ethics, social implications, and other deeply reflective ways of thinking about data science." Different fields have different characteristics and different needs: it is important to recognize the differences between social sciences, humanities, arts, professional schools, natural sciences, engineering. A diverse ecosystem is seen here as incredibly valuable and important, as well as having undergraduates involved.

III. Open Q&A on high-level themes:

The panelists offered interesting responses to several questions posed by the audience, which further framed their views on DS program development within their respective institutional contexts. These perspectives did not coalesce into a single path; in fact, they underscore how institutional differences helped to propel implementation and growth. These differences include, for example, focus on undergraduate versus graduate students; programmatic funding through a single entity vs. campus-wide monetary stimulation or incentive; advances to the conceptualization of Data Science and "intellectual ownership."

Questions from the audience concerned metrics, and assessment of interdisciplinary "uptake" or work in Data Science; understanding how to manage "turf" concerns as Data Science takes takes hold and grows, and relatedly, avoiding new silos; maintaining faculty engagement, especially where there is concern of future administrative or T&P conflicts for those with joint appointments; and, making sure that structures, functions, and processes are designed to be inclusive, and that engagement, participation, and outcomes are evident of this.

Question: How to tease out what participation is? Who is active in the data science community? How is this measured at your institution? Who is the driving force?

- **Jeannette Wing:** Adds: How are you, at your universities, thinking about it? Responses to the survey conducted as a result of the inaugural Academic Leadership Summit report showed all levels of granularity, from machine learning to sub-disciplines, to humanities, broadly.
- **Brian Athey:** We are flooded with data and we need to make decisions based on this data. In this way, data science can be deeply transformational. We feel that we need methodological tools and domains coming together to have the maximum impact.
- **Cathryn Carson:** As David Culler said, it's not a tree, it's a graph. Yes, computer science and statistics have to be central at Berkeley -- and yet, they are not responsible for the whole.
- **Melissa Cragin:** Restating Tyler's question: How can we get some quantitative measures of interdisciplinarity in data science?

Question: How do you manage the intricacies of dealing with other institutes on campus (e.g., AI institutes) if you have them? There are concerns over ownership of activities and topics.

- **Saul Perlmutter:** Data science entities should acknowledge the value of other institutes on campus, but it's hard to imagine that any institute on campus is serving all the needs. Data science programs should find what needs aren't being met and help meet them.
- **David Culler:** We didn't have a lot of money to start. Meanwhile: Who owns the intellectual agenda? This is a bad question. Who cares? We shouldn't, as academics, be

asking the question: if I own the agenda, do I get more TA dollars? Academia moves slower than everyone else.
- **Brant Cheikes:** At UMass Amherst, there is little inter-institutional conflict. There are good relationships and we have strong collaborative ties to other institutes. There is no fighting over the intellectual agenda. The Center is well-resourced, compared to other schools and colleges on campus. We try to keep things friendly by doing favors for one another. In particular, we share computational resources for strategic and natural collaborations. In general, we find ways to help each other.

Question: How do you go about creating cross-cutting data science entities (and not just another silo)? How do you get the attention of the administration to have them get behind a new paradigm for data science programs? What kind of leverage can you have? How does the reporting structure relate to autonomy of the entity?
- **Magda Balazinska** When we created the eScience institute, everyone was invited. Anyone who said they were interested in what we were doing, we said “great!”. If they say, “I want my department to run it," we say, “Okay, eScience is running it and you are eScience”. We allow anyone who is interested to have a say in frameworks. So, if they complain, we say, “We made this together! We can work together to make it better.” If they have a great class they teach, we integrate it into the framework.

Question: How do we promote engagement among faculty members? How do you decide who gets to be formally involved in the group (without being a club)? How do you decide when someone has fallen away?
- **Brant Cheikes:** We have thought about this a lot! Our faculty engage with and contribute to the Center to the extent that they see the Center supporting their career success. We have found ways to do this over time and have steadily expanded the cadre of faculty who truly see themselves as Center faculty. One area where we can do better is promoting ourselves to the broader faculty. Some faculty are disengaged due to lack of awareness of what the Center does and how it can help them.
- **Brian Athey:** This requires fluid and practical decisions. We see a call from DOD and then form a working group, have lunch, and see where things go.
- **David Culler:** It’s a give-and-take. Be explicit about sharing (e.g., MOUs) in a structured way. Expectations and explicit discussions help build trust and stronger connections. If you think about building credit into the structure of these relationships, you are fighting less.

Question: With shared faculty lines, how do you sustain engagement with the data science program when there is a conflict between the home departments and the data science program (over tenure, for example)?
- **Julia Kempe:** We hire faculty that are young and engaged, and they feel like they’re building something new. Tenure is decided between the Center and the home department, so service will be counted. Be explicit in MOUs about distribution of tenure weight. There is a lot of goodwill.

Question: How do we organizationally ensure that we build representative students and faculty? What organizational decisions ensure inclusion in the data science community?
- **Cathryn Carson:** Note that if you start at graduate level, you are already dealing with a self-selected population.
- **David Culler:** For Berkeley, Data 8 enrollment is 52% is female, and about 12% are underrepresented. Meanwhile, there is a tremendous pipeline issue for faculty. The postdoctoral and graduate pipeline is easier to work with, in this regard.

Observation: Interestingly, a question about measuring engagement (or participation), and impact on interdisciplinary arrangements went largely unanswered.

Acknowledgements
The 2018 Data Science Leadership Summit was made possible through generous support from the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation. This report was completed with contributions from others, including AAAS Fellows who attended the meeting and assisted with note-taking. The authors also thank the Panel co-Chair and Program Committee member, Ed Lazowska, as well as the panelists, for the making of such a successful session, and for reviewing this report. It is stronger for their contributions.

Melissa Cragin San Diego Supercomputer Center University of California, San Diego mcragin@sdsc.edu	Tyler Kloefkorn Board on Mathematical Sciences and Analytics National Academies of Sciences, Engineering and Medicine

References

Jeannette M. Wing, Vandana P. Janeja, Tyler Kloefkorn, and Lucy C. Erickson. 2018. Data Science Leadership Summit: Summary Report. Technical Report. National Science Foundation, USA. https://dl.acm.org/doi/book/10.5555/3293458

ABT Associates. 2018. Academic Data Science Centers in the United States. https://doi.org/10.6069/H63V-V078

The National Academies Press. 2018. Data Science for Undergraduates: Opportunities and Options. https://www.nap.edu/catalog/25104/data-science-for-undergraduates-opportunities-and-options

The Moore-Sloan Data Science Environments. 2018. Creating Institutional Change in Data Science. https://doi.org/10.6069/v5s4-8n41