UC DAVIS
UNIVERSITY OF CALIFORNIA

DATA SCIENCES INITIATIVE

IMPLEMENTATION COMMITTEE REPORT

SUBMITTED TO PROVOST HEXTER

MARCH 18, 2014

I. EXECUTIVE SUMMARY

The Data Sciences Initiative (DSI) Implementation Committee was established in October 2013 with the goal of defining an action plan for the initial phase of the UC Davis Data Sciences Initiative. The committee met ten times over the course of the last five months to explore and frame the initial intellectual, technological, staffing, infrastructure and operational elements needed to support the data science work of UC Davis faculty, researchers and students. Building on the vision articulated last year by the Big Data Implementation Committee, and informed by additional research and consultation both on and off campus, the DSI Committee has developed a plan with both strategic and programmatic elements:

- *Strategic objectives* – Four objectives, each with a complementary role in establishing a strong foundation for the data science program at UC Davis, will guide the work of the Data Science Initiative in its inaugural year;
- *Infrastructure needs* – Campus and cloud computing, large-scale storage and data analysis capabilities, along with reliable high-speed networking, visualization display and other technologies will be critical to the success of the DSI. These needs will be further defined through consultation with faculty and researchers;
- *Space needs* – To effectively fulfill its research, education and outreach mission, the DSI will require appropriate office space for faculty, staff, graduate students, and visiting scholars as well as rooms to hold conferences, seminars and other DSI research and teaching functions. Space has been secured in Shields Library for this purpose; and
- *Staffing needs* – A small core team of DSI leadership and staff, with the appropriate expertise and skills, will be necessary to carry out the objectives of the DSI's first year. These appointments are a top priority for the Initiative.

The action plan presented in this report is intended as a roadmap for the DSI Founding Director to refine and implement during the 2014-15 academic year, in close consultation with the University Librarian and the DSI Advisory Committee. Until the DSI Advisory Committee is fully established, we recommend that the DSI Implementation Committee continue to serve in an advisory capacity to the Founding Director and help ensure the Data Sciences Initiative is off to a strong start. We further recommend that at the end of the first year, the Founding Director submit a brief report to the Provost summarizing the first year's activities and progress, along with plans and funding needs to support Year Two objectives.

II. DSI: OBJECTIVES FOR YEAR 1

The mission of the UC Davis Data Sciences Initiative (DSI) is to support and enable Big Data science to accelerate discovery at the frontiers of scientific, engineering and social disciplines. Keys to its success will be the scientific excellence of its leadership and core members, its visibility on campus, education and training of students, as well as its ability to identify and meet the needs of the UC Davis community. To maximize its impact in research, education, and infrastructure, we therefore recommend that the DSI reach the following four objectives during its first year.

1) Build the DSI core human resources

We recommend the following appointments:

- ***An Associate Director for Education***. It is in the core mission of the DSI to assume leadership in data science education at all levels on campus, from undergraduate students to faculty, including graduate students, postdoctoral fellows, and staff. The role of the Associate Director will therefore be to define an initial data science education strategy n, as well as to develop marketing plans for existing educational programs on campus related to data sciences and data analytics. We anticipate that this appointment will be a current member of the UC Davis faculty who will commit to spending a portion of his or her time developing the educational programs.
- ***An Academic Coordinator.*** The main role of the academic coordinator will be to identify the key services that can be provided by the DSI, as well as to define the budget, staffing, and assessment strategies for these key services in order to maximize their impact on campus.
- ***Members of the Internal Advisory Committee.*** The Internal Advisory Committee (IAC) will assist the Director and Associate Director by reviewing the progress of the DSI and by providing advice and guidance on how to best utilize campus resources, such as the Center for Information Technology Research in the Interest of Society (CITRIS) and the UC Davis Clinical and Translational Science Center (CTSC) among others. The IAC will also serve as a key support group for the DSI by linking to UC system-wide efforts and resources, and by working with the Director to identify new opportunities, including external funding opportunities.
- ***Members of the External Advisory Committee***. The External Advisory Committee (EAC) will assist in the guidance and direction of the DSI by providing annual written recommendations to ensure the cohesion of the program. The members of this committee can be recruited from industry, academia, governmental agencies, and thoughtful leaders. The committee will identify ways in which the DSI may benefit from resources at other universities and industry. The EAC will include experts in relevant areas of data science.

In addition to making these initial appointments, we recommend that **domain experts** be identified that will serve as liaisons to the different colleges and programs on campus. Finally, we recommend that a strategy and timeline be put in place to help identify and **hire a permanent director** for the DSI.

2) Develop a strategic plan for the DSI

To guarantee institutional impact across schools and colleges on campus, the DSI will require an understanding of the needs of the faculty related to data sciences, a plan to provide solutions to these needs, a vision for opportunities for the data sciences in teaching and research, and a strategic plan to guide the implementation of the DSI. The strategic plan should identify the needs and opportunities on campus as well as potential key external and campus partners for the DSI. A key objective for the first year will be to establish the governance for the DSI to ensure that it can effectively reach its goals.

3) Develop the physical space for the DSI

The DSI is set to provide an intellectually stimulating physical environment to nurture collaboration between theoreticians and scientists who tackle the challenges of Big Data, and offer a space for domain scientists to interact with these experts. A suitable space has been identified on the third floor of the Shields Library. We recommend that a committee be formed to help the Director and Assistant Director renovate and set up this space such that it meets the needs of the DSI. Special consideration must be given to its IT infrastructure and technology resources , as the DSI is expected to serve as an antenna for data science technologies for the entire UC Davis campus.

4) Increase the visibility of the data sciences and the DSI

The academic coordinator, with help from the Director and Assistant Director - will work with programmers, developers, and system administration staff to provide an integrated web-based portal to facilitate collaboration as well as access to data and tools by both DSI affiliates and members of the UC Davis community. Creation of a central community-based portal is driven by the principle that the DSI needs to both support and advocate for an accessible environment where a vast spectrum of investigators can find information pertinent to their interests in Big Data analytics as well as tools, and potential partners involved in translational research. When fully developed, the central repository will allow researchers to browse tools to advance their work, while developers will be able to access and contribute to source code that will be available for these tools.

III. KEY ELEMENTS OF THE UC DAVIS DATA SCIENCES PROGRAM

a. Staffing

Beyond its part-time Founding Director, the Data Sciences Initiative will require dedicated staff to carry forward its initial agenda. We intend the DSI's staff to be minimal and strategic, leveraging the capabilities of current faculty and staff across the university and beyond as much as possible.

A key position to launch the Initiative will be a full-time Academic Coordinator to support the Founding Director and serve as the operational lead. The Coordinator will develop approaches for outreach to current faculty, identifying and exploring existing services and expertise relevant to Data Science. The Academic Coordinator will also manage the space re-development, help with grant writing, managing the budget, as well as the myriad other aspects of running a large and complex project.

Given the Initiative's emphasis on building an educational program in data sciences in partnership with the UC Davis schools and colleges, we also anticipate the need for a part-time Associate Director for Education. The Founding Director places a high priority on the Initiative's educational programs; he will be focused however on developing its strategy and its outreach across campus. An Associate Director for Education is therefore expected to provide the complementary and necessary resources to identify current courses, develop new courses or data sciences modules, coordinate with existing educational efforts on campus that help

students become more quantitative, and coordinate the DSI efforts in education with the Academic Senate. The new courses developed as part of DSI will be aimed at students across the undergraduate to post-graduate spectrum in a variety of disciplines with differing expertise and expectations for data science.

The Data Sciences Initiative is expected to contribute to and develop externally-funded grants to implement its vision and augment its budget. To help with administrative tasks including budget, grants, and personnel management, the DSI will provide funding for administrative support to the University Library to supplement its existing staff. The Library has a full complement of administrative services, but its operational profile differs from the Initiative so modest additional resources will ensure that the Initiative is supported appropriately. In particular, the Library has administrative staff to help with the space renovations, human resources procedures, and day-to-day procurement transactions, but has very few resources to support grant management and work with extramural funding. There are currently staff within the Library who can take on this additional charge; they will require back-filling to continue to manage the Library's normal operations.

Since the Data Sciences Initiative is envisioned as a bottom-up operation that will develop as campus needs and opportunities are more clearly defined and as we discover what services and expertise related to data science exist on campus today and where there are gaps, the staffing model beyond these key positions will be identified over the first year.

We anticipate wanting to recruit a few faculty members or other researchers currently on campus to participate actively in the work of the Initiative and may need to provide funds to buyout their course load or offer stipends. We further anticipate the need for staff to complement existing expertise on campus but would not speculate at this time on exactly what skill sets are needed, or which model we will implement to make these skills available to the campus community at large. Possible models include but are not limited to creating new core services directly managed by the DSI or providing additional support to existing cores to increase their areas of expertise and improve their outreach. The expertise we may need to acquire includes such areas as database management, scientific computing, programming in specialized languages, knowledge of specific applications or computing hardware, text or data mining expertise, statistics, and machine learning. Many of these skills are already available on campus but we need to learn how widely they are available beyond their home departments and how well they match the consulting, education, and research partnership needs we identify in Year 1.

The Data Sciences Initiative is currently funded for two years so that its Founding Director may identify the really ground-breaking opportunities for UC Davis in the data sciences and develop detailed plans for how to address them. We expect the Initiative to evolve into something much larger and transformative to campus – an Institute or similar major research and education focused unit – and for that reason we anticipate needing a permanent, full-time Director with significant expertise and leadership experience in data science to carry our work forward. We recommend the campus begin the corresponding recruitment process early in Year 2 since it is a competitive field and finding the right person may require significant outreach.

b. Partnerships

The Data Sciences Initiative should serve as an important bridge between UC Davis and potential partners. Partnerships through data exchanges of various scales and formats, and for a

wide variety of possible purposes, provide opportunities for Davis faculty, staff and students to interact with peers and collaborators at other institutions, and influence directions in data sciences beyond campus. In addition to research collaborations, the emergence of powerful technologies enabling distance learning techniques opens up a broad range of possibilities that DSI may choose to explore. Opportunities to access hardware and software capabilities of others should certainly be explored, to determine whether this may be a cost-effective route to partially defray the costs of the cyberinfrastructure investment needed at UC Davis. The possible solutions worth considering range from purely commercial contracts with corporations such as Amazon, Google, Microsoft or IBM, to mixed service/joint R&D partnerships with national laboratories. Campus partnerships with 3 national labs (Lawrence Livermore, Lawrence Berkeley, and Sandia Livermore) are at various stages of development and include cyberinfrastructure components. Other possible partnerships for consideration might be with the UC San Diego Supercomputing Center, NASA Ames or the National Center for Supercomputing Applications at Illinois.

Collaborative R&D opportunities with other universities, corporations and national labs will be greatly enhanced by the foundation of the UC Davis Data Sciences Initiative. As a matter of priority, DSI needs UC Davis to invest in a computing test-bed structure to enable the evaluation of the viability of partnering with others, be it on a commercial or collaborative basis.

c. **Space**

A key component of the Data Sciences Initiative's initial report was to create dedicated space to allow the co-location of experts from across campus and to bring together researchers with data needs with the people who can help them. The report recommended that the Initiative implement a "hotel model" where guest researchers could be co-located with the Initiative's experts for a period of time, leading to much greater exchanges of ideas and information than is possible in remote, distributed, and virtual arrangements.

To achieve those goals and house the Data Sciences Initiative, the University Library has generously provided space in Shields Library that was recently turned over by the Institute for Government Affairs ORU. The space is ideal for this phase of the Initiative in that it has a combination of private offices, cubicles, open spaces, and a 50-seat lecture area.

The space has not been improved in many years and would greatly benefit from renovation, such as upgrades to the wired and wireless networks, better lighting, more power, new furniture, and so on. IET staff experts are currently assessing the technology infrastructure, and Library facilities staff has begun getting estimates for basic improvements (cleaning, paint, etc.)

Over the summer the Founding Director will work with Library staff and Administrative and Resource Management staff to determine the appropriate modifications to the space and develop a detailed budget for review and approval by the Provost's office and ARM leadership. A budget of up to $500,000 has been set aside in the Institute's start- up funding. Every effort will be made to preserve some of that funding for future improvements as the staff come into the space.

d. **Infrastructure**

The IT infrastructure directly serving the DSI is expected to encompass high-speed networking, collaborative technologies such as video conferencing, and small-scale compute, storage and visualization technology to support web sites, portals, demonstration projects, solutions development and educational activities. Beyond demonstration and initial development activities, the DSI will work with the campus community to locate and leverage existing IT infrastructure and services, both on and off campus, that best address translational research and educational needs.

Data center services will be a core element necessary to the success of data-driven science on campus. Most activities of the DSI (both educational and research) will require housing of servers, storage and the use of high-performance computing (HPC) at a data center due to space, power, reliability and security needs. Some progress has been made to develop an HPC program at UC Davis and continued investments in HPC will provide infrastructure and services critical to the DSI. While the DSI will not lead the development of the HPC program, close coordination between the DSI and the HPC program will be important to the DSI's goal of promoting and coordinating services for data sciences. The DSI and the HPC program are expected to leverage the proposed joint data center infrastructure; and in fact, they further underscore the need for it. The existence of a modern data center and the DSI's use of its facilities and services would promote data science programs and attract members to the DSI.

e. **Budget/Funding**

The initial budget for the Initiative is $1,500,000 over two years, including the $500,000 that is earmarked for the space improvements. The $1,000,000 aimed at the Initiative's work will be budgeted for Year 1 to begin on May 1, 2014, and Year 2 on May 1, 2015 (or more realistically, July 1, 2015).

The Founding Director will work with the University Librarian to develop the Year 1 budget of up to $500,000, in consultation with the campus deans and relevant vice provosts (e.g., the Vice Provost for IET and CIO), consistent with the Provost's August 12, 2013 "Response to the Big Data Implementation Report" letter. The first year budget will include the necessary staff (approximately $180,000 for the Academic Coordinator, Associate Director for Education, and added administrative support) and three funding pools for:

- Course buy-outs of faculty and instructors needed to teach new data science courses ($60,000);
- Undergraduate student Research Assistants and Master's students to provide low-level expertise as part of a core service ($50,000); and
- Recharge costs to acquire expertise from existing staff around the campus, e.g., Bill Broadley for HPC, IET for systems administration, Statistics Lab staff, etc. ($140,000).

The budget will also set aside $20,000 to commission a website as part of the Initiatives' core service offering, which will then be maintained by DSI staff.

These combined budget items total $450,000, and we will defer the remaining $50,000 to supplement the Year Two budget as the more detailed staffing and space needs emerge.

A Year Two budget proposal will be submitted in the spring of 2015 for the Provost's approval , depending on the Initiative's progress to date. The budget for the Initiative will come as a pass-through from the University Library and for the duration of this arrangement will be a separate line funded by the campus. The Founding Director has full authority over the budget and is responsible for all administrative decisions, including management and performance reviews of additional Initiative staff.

If the Data Sciences Initiative attracts extramural grant funding directly, it will be administered by the Library and the departmental share of F&A costs will be distributed to the Library, consistent with the campus budget model, for re-investment in the Initiative.

By the end of its first two years of existence the Initiative will need a permanent budget to support its ongoing operations. This budget is expected to be primarily funded by the deans of the relevant schools and colleges, with only minimal central campus funding. The long-term budget model will also be reviewed at the end of the first year and before releasing the second year funding, to ensure that a trajectory toward a sustainable funding plan has been achieved. Should the Initiative have a satisfactory long-term budget model at the end of two years but require additional one-time bridge funding to reach full sustainability, the Initiative's leadership may request that funding as part of the FY17 budget planning process.

Part of its ongoing sustainable funding model will necessarily involve research grants to staff or affiliates of the Initiative. Given the current model for F&A costs returning to the 'home' department of the grant and the need to incentivize faculty from departments across campus to participate in grant projects, we will develop a fairly sophisticated cost sharing model for the F&A returns so that both the department and the Initiative gain from participating in the project.

f. Governance and Membership

To ensure its utility and success as a campus-wide initiative, gain maximum support from the campus research community, and to help develop its long-term sustainable funding model, the Data Sciences Initiative will need a transparent and inclusive governance model.

While the Initiative is initially funded by campus and housed in the Library, to achieve its long-term vision and be transformational to UC Davis, the governance of the Initiative must go far beyond the Provost's office. Governance in this context has three aspects:

- An internal, administrative steering committee or similar, consisting of key deans, faculty members, and other administrators (e.g., from the Office of Research) who have a stake in the success of the Initiative. This group will have responsibility for working with the Founding Director on the Initiative's mission and initial strategic plan;

- An external, scientific steering committee or similar, with expert membership drawn from across campus and beyond. This group will further identify current faculty and academic programs relevant to the goals of the Initiative, and form collaborative partnerships in furtherance of the DSI strategic plan; and

- A 'membership' model for participating faculty and students that includes who can become a 'member', by what process, with what benefits, expectations and assessments. Core faculty members identified through this process will work with Initiative staff and

other campus stakeholders to identify core infrastructure needs, including technology, space, and staffing, to achieve the goals of the DSI as articulated in the DSI strategic plan.

We propose to create an ad hoc advisory group for the period of May through September, to assist the Director with short-term planning and help to get the longer-term governance structure in place. We will work with the current Implementation Committee to identify members for the ad hoc group before we disband later this month.

The two oversight groups will be identified and charged by the beginning of the 2014-15 academic year. Members of the current Implementation Committee and the original Big Data committee will be encouraged to participate in one or more of these groups, and the Deans will be invited to suggest how they would like to be involved on an ongoing basis.