# A Case for a College of Computer and Data Sciences (CCDS)

Version 1.0, J Sventek, 20 May 2019

## 1 Executive Summary

As computing has become pervasive throughout our universities, many AAU and other R1 institutions have created colleges/schools of computing to facilitate interdisciplinary curricula and research. The deliberations of the Data Science visioning committee during academic 2017/18 recommended the creation of a Data Science Program, organizationally outside of any currently existing UO College/School, to facilitate interdisciplinary curricula and research in data science; this program would have core faculty and affiliated faculty from other units across the university; most faculty in the Department of Computer and Information Science will be affiliated with such a program. This white paper argues for the creation of a College of Computer and Data Sciences (CCDS) to provide an organizational structure to facilitate world-class computer and data science curricula and research.

## 2 Introduction

There is a growing trend in transforming a department of computer science, usually housed within a college of engineering, science, or arts & science, into a school/college of computing. Additionally, this transformation is often accompanied by inclusion of departments of information systems (iSchools), Information Technology, Computer Engineering, and other related units. This trend is designed to facilitate interdisciplinary, computational curricula and research across the university.

A visioning committee during the 2017/18 academic year advised the provost regarding the organizational structure needed to fully animate the Presidential Data Science Initiative. The task force recommended the creation of a Data Science Program, similar to the Environmental Studies Program in CAS. The interdisciplinary nature of the Data Science Program required that it ***not*** be homed in any of the ***currently existing*** colleges or schools. The program is to use a hub and spoke structure, with faculty in the hub having the program as their tenure home and faculty in the spokes being affiliated faculty from other parts of the university. Some of the hub expertise, such as machine learning competence, would also come from strongly affiliated faculty (e.g., CAS/CIS).

The increasing trend to create schools/colleges of computing, the need for the UO Data Science Program to be outside of any currently existing college/school in order to fulfil its interdisciplinary mission, and the strong affiliation of CIS faculty with the program argue strongly for the creation of a College of Computer and Data Sciences (CCDS) at the University of Oregon.

## 3 Schools/Colleges of Computing[1]

As computing continues to grow rapidly and to permeate universities' intellectual landscape, many departments find that their programs have outgrown, or are outgrowing, the confines of their current organizational structure, typically as a department in colleges of engineering, science, or arts & sciences.

[1] The material in this section is liberally taken from https://cra.org/resources/creating-institutional-homes-for-computing/. The white paper summarizes the discussions from two successful panels at the 2016 and 2018 biennial Snowbird Conferences on this topic.

Such findings lead to exploration of expanding into a school/college of computing. Such a move can lead to significantly improved interdisciplinary opportunities.

In the following, the term “computer science” refers to the core disciplines concerned with the theory and design of computer systems; it includes such topics as algorithms, programming languages, artificial intelligence, operating systems, networking, databases, distributed systems, human-computer interaction, and software engineering. The term “computing” includes computer science, as well as its broader applications, including robotics, graphics, interaction design, language technology, information systems, and data science. Neither term is to be interpreted as being precise, but should give a distinction between core computer science and computing.

## 3.1 Why become a School/College?

Computer science is different from other natural science departments:

- Computer Science is a “Science of the Artificial” – it focuses on how to orchestrate the internal theories, techniques, and tools of the discipline to address issues in the natural world. Central to this orchestration is a science of design – i.e. the ability to map artificial concepts to natural world issues.[2]
- Armed with this ability to design at the artificial/natural boundary, a BS in Computer Science degree leads to a wide range of high-paying job opportunities immediately upon graduation related to completed course work; students also have well-paid internship opportunities. Such opportunities for BS majors/graduates in many other natural science disciplines are less plentiful.
- Faculty (and PhD graduates) have a wide range of opportunities outside of academia; in many sub-areas of Computer science, faculty are actively recruited by industry. Non-academic job opportunities for faculty in other natural science disciplines are usually less commonplace.
- Increasingly, professional degrees in specialized computing fields (e.g., security, data science) are being created, with significant enrollment. This is not such a significant differentiation at UO given successes in the Professional Internship degree programs currently housed in the Knight Campus.
- Besides being relevant to the natural sciences, computing and data science are relevant to almost every area and domain outside the natural sciences.

Computing has become pervasive through our universities and colleges; nearly every discipline needs computing. The case can be made that computing should be its own school/college on a campus to manage these needs throughout the campus. A college recognizes the role computing plays in all disciplines, in research as well as academic programs, and in job opportunities for non-computer science majors. In a small number of cases a single department (computer science) morphs into a school/college; more typically, the school/college is created from computer science, information sciences, statistics, computational biology, digital media and art, MIS, information technology, and communications (where such units already exist).

Schools/colleges of computing provide excellent opportunities for broad, multidisciplinary coverage and increased research funding. A school/college of computing is generally more inclined to hire faculty from a wide range of academic disciplines who are doing research related to computing. In addition, many

[2] Herbert A Simon, The Sciences of the Artificial, The MIT Press, Cambridge, MA, 1996, ISBN 0262193744.

current funding opportunities build on all areas related to computing; a school/college of computing is in a much better position to respond to such funding opportunities.

There are excellent job opportunities for students pursuing the wider set of degrees in schools/colleges of computing. Job demand is high for both traditional computer science graduates as well as more application-oriented students.

Schools/colleges of computing provide excellent platforms for external fundraising. The style and inclinations of computing/IT-related people, who are generally the primary donor prospects for schools/colleges of computing, are quite different from those of donors from engineering or general business careers. Thus, a dean of a school/college of computing may be considerably more successful in relating to these computing-related donors than a dean of a more general school/college.

The "direct entry" of students into CS-related employment upon graduation argues for a strong industry relationship in terms of advisory boards, collaborative research, internship opportunities, and other aspects. It is easier for a more focused College, such as CCDS, to establish and maintain such industrial relationships to the benefit of students, curricula, and research.

As computer science has become a popular major, and the need for computing has reached every college on campus, enrollments are booming nationwide. Managing such booming enrollments and the changing needs of computing instruction require significant resources. A school/college is in a better position to make the case for additional resources relative to a department in a larger college structure.

A table of universities that have created Schools/Colleges of Computing may be found at https://cra.org/resources/creating-institutional-homes-for-computing/.

## 3.2 Challenges

The current home for a computer science department may be concerned about the loss of funding, academic synergies, potential harm to industrial relations, and fundraising if computer science becomes part of a new school/college. The university administration needs to be convinced that creating a school/college is in the best interests of the institution, students, and other stakeholders.

It will be necessary to form departments once the school/college reaches a significant faculty size or if the school/college is created from two or more already existing departmental units. It is desirable to keep barriers between departments low to facilitate shared teaching, student supervision, and joint research. Additionally, a primary reason for creating the college is to facilitate joint research with other colleges to maintain the interdisciplinary vigor of computing.

Is the "brand" the department or the school/college? Individual faculty generally have their current home department as their brand identity; a school/college may want faculty to broadcast the larger unit as their identity. There needs to be some of both.

Schools/colleges are expected to perform many more functions than a typical department. The necessary resources, personnel, and expertise to take on these functions (e.g., development, career planning, student support, and communications) must be acquired.

# 4 Data Science Program[3]

The provost charged a Data Science Visioning Committee to generate a proposal that addressed four major categories of consideration for the creation of a data science program: tools and technical resources, space requirements, educational framework, and initial intellectual foci. The committee delivered the results of its deliberations in the document entitled "A Vision for Implementing the Presidential Initiative in Data Science at the University of Oregon".

## 4.1 Overall principles

The committee agreed upon several guiding principles for the implementation of the initiative that would lead to the highest probability of success. These principles were informed by advancements in the field of data science in general, development of data science programs at other academic institutions, and needs for data science in society. The principles also build upon UO's history of research and educational excellence and existing efforts in data science, while also looking to possible areas of growth and impact. These principles informed the committee's specific guidance (below) for structuring the implementation of the initiative. The subsequent bullets enumerate those principles that are relevant to the topic of this white paper.

The data science initiative should…

- *Be broad and well connected across the university.* Because data science impacts nearly all areas of the university and society, the initiative should encompass the diversity of different units on campus that are willing to make substantive contributions. The initiative should have focal areas in each unit on campus, but **it should not be sequestered in any one existing school or college**. A successful data science initiative will form a network of collaboration across campus and beyond.
- *Include both research and education.* Adding novel areas of scholarship and discovery in data science will be a crucial component of a successful initiative, but so will training the next generation of data scientists in nearly all methodology and domain areas. While some aspects of education in data science already exist at UO, more can be done to provide education in both core methodologies and domain applications of data science.
- *Build upon the long history of interdisciplinary research at UO.* A long history of highly successful interdisciplinary research exists across campus because of the size of our university, and the lack of traditional programs in (for example) engineering and medicine. These interdisciplinary efforts have been supported in a variety of ways across campus. Schools and colleges, departments and institutes, and centers and programs have all brought together scholars from diverse backgrounds. Interdisciplinary research has been an engine for increasing collaborative research, support, and contributions that have significantly impacted the world. Incorporating this interdisciplinary ethos into the data science initiative will serve it well.
- *Capitalize on the identity and strength of UO being a liberal arts research university.* In addition to the natural sciences and professional schools, the social sciences and humanities have a long history of scholarship and educational excellence at the university. A modern data science initiative will also be built upon this strong liberal arts foundation. Doing so will allow data science to become an additional component of a liberal arts education, the historical goal of which is to provide students with a broad education to become successful in the world. It would

[3] The material in this section is liberally taken from "A Vision for Implementing the Presidential Initiative in Data Science at the University of Oregon", https://provost.uoregon.edu/files/data_science_report.2018.docx.

also encourage the full inclusion of liberal arts knowledge and scholarship into the data science revolution in ways that would be uniquely University of Oregon.

- *Leapfrog historical structural weaknesses.* A negative consequence of the lack of schools of medicine, engineering and agriculture at UO has meant the university has concentrated less effort on building in areas such as applied math, statistics and computer science. While all of these areas exist at UO, they are relatively smaller in comparison to our peer institutions. A reason has been the appropriate focus on strength in 'pure' areas in the departments of mathematics and computer and information sciences. As a consequence, many of the math and computer science application areas have grown within domain areas.
- *Advance new research and educational opportunities.* Several areas of scholarship and research in data science are underrepresented at UO, and a successful initiative will increase these areas. Similarly, there are segments of the undergraduate and graduate student population that are not being fully served by present course offerings in data science methodology, or data science applied to domain areas.
- *Help build bridges to new academic and societal partners.* A successful data science initiative should act as a portal for increased impact of UO on society. It should also act as a conduit for new research and educational opportunities to flow back into the university. The initiative should focus on building collaborative efforts with researchers at other academic institutions, particularly those within the state (e.g. OSU and OHSU) as well as along the West Coast. In addition, the initiative should increase the impact of UO scholarship and education on society by forming key relationships with key economic and governmental partners, as well as other segments of society.

## 4.2 A vision for the overall data science ecosystem at UO

The principles lead to proposed structures to support the data science ecosystem at UO. The term ecosystem is intentionally used, because the vision of a successful data science initiative comprises numerous interacting and interconnected components both within and outside the university. The committee developed a 'hub-and-spoke' model for the overall ecosystem (see figure), which attempts to capture these connections. The specific designations for each of the components are not meant to be restrictive or exclusive, but are simply meant to provide real context for the overall conceptualization of the data science initiative. The goal of the hub-and-spoke ecosystem model is to support and promote connectivity along existing areas of scholarship and education, but also to support and guide the organic growth of the initiative.

Because of the overall breadth of data science across the university, and its connection with partners in the wider world, the proposed ecosystem vision consists of three major categories. Moving from the inside out are the following major categories (see figure).

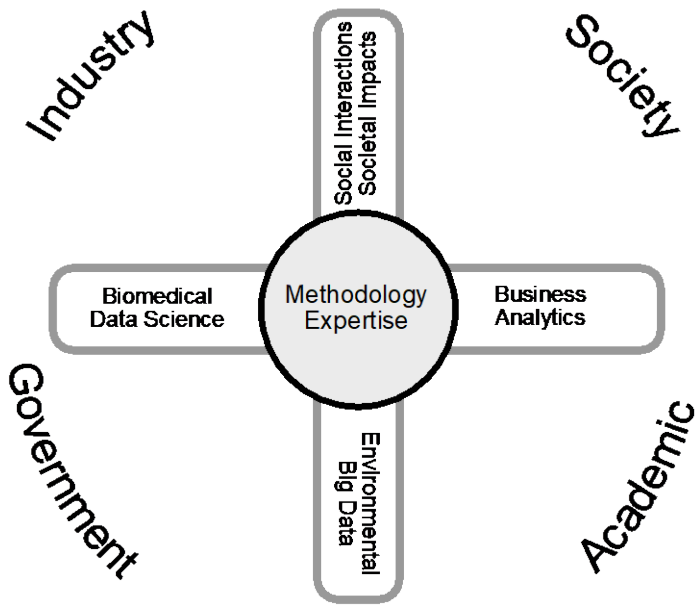



Diagram of hub and spoke for UO

- Methodological hub. This is the area which contains much of the research and educational expertise for implementation of data science. Many of the technical aspects of data science, such as data collection, database creation and curation, as well as statistical and computational analysis, would be considered components of the hub. Because of the need to effectively present and communicate data, as well as make ethical and legal decisions regarding the appropriate use of these data, several aspects of design, communication, ethics and law could also be components of the core. A key aspect of the contributors affiliated with the hub is that they can pivot to work with many different spokes either through research collaborations or educational activities. For example, faculty who focus on machine learning or data science ethics can collaborate with many different domain data scientists and teach core competencies to students drawn from the different domains. Some examples of areas of inquiry within the hub might include but are not limited to:
    - Applied statistics. Recent prominent machine-learning applications have underscored the point that access to "big data" does not substitute the need for statistics. Even given voluminous datasets, it is crucial to consider statistical issues such as dimensionality, (model) over-fitting, and causality. New statistical learning and dimensionality-reduction techniques are essential for the continuing evolution of data-science applications.
    - Data management and cyber security …
    - Machine learning and artificial intelligence …
    - Data science interpretation, visualization and communication …
    - Data ethics, law, and policy: issues in responsible research practice but more significantly research focused on forecasting and mapping the social and ethical ramifications (both positive and negative) of emerging data science techniques, especially impacts that are unanticipated.

- Domain spokes. These are areas where the research and educational activities are focused primarily on a particular field or problem (e.g. biomedical data science or business analytics). Faculty in these spokes may gather large amounts of data and use advanced analytic tools that are common to those in the hub, but their primary research and educational focus will always be first on their domain area of expertise. They can therefore collaborate with individuals in the hub and have connections and impact outside of UO, but will primarily do so through their fields. Importantly, similar to how research institutes at UO cross traditional departmental and school boundaries, so too can the spokes, which may bring together researchers even from different colleges. Below are four examples of spokes that describe related research activities that are occurring at UO and which are natural avenues for focus and expansion. Some initial spokes might include but are not limited to:
    - Business analytics. Industry is keenly interested in leveraging data to identify new investment opportunities and improve efficiency. As firms become increasingly adept at collecting and structuring the data surrounding their operations, understanding how to exploit that data becomes increasingly relevant. The Lundquist College of Business has hired several new faculty with research focused on data-analytics methodologies—including applications within the health care, finance, and logistics industries—and continues to grow its faculty ranks and courses in this critical area.
    - Environmental big data analytics. …
    - Data science of social interactions and societal impacts. …
    - Biomedical data science. …

- Ring of connectivity. While the hub and spokes are largely (but not exclusively) located within the university, the goal is to have greater impacts on the world by increasing connectivity with other academic institutions, government agencies, and industry partners. The overall structure of the data science ecosystem allows for the increased research and educational connectivity both through the hub as well as around the ring, allowing feedback to accelerate novel research and education within our university informed strongly by the wider world.
    - Academic partners. … brief example for each …
    - Government agencies. …
    - Industry partners. …
    - Non-governmental and non-profit partners. …

## 4.3 Vision for the initiative to be an engine for expanded research

For the data science initiative to fulfill the goal of supporting advanced interdisciplinary research, it will require building the components of other successful research departments and institutes at UO. These include the development of intellectual support networks such as shared physical spaces and computational resources, as well as activities such as seminar series and journal clubs. In addition, administrative support for developing, submitting and managing grant proposals will be required. For the initiative to be maximally successful it will also develop mechanisms for incubating novel research ideas and directions.

The most important component of a successful research institution or department involves the involvement of faculty, students, and staff. For the initiative to succeed as an initiative, it should develop various degrees of relationship with the program and commensurate expectations.

For tenure track faculty (TTF), these affiliations could be the following:

- Affiliation of existing or new faculty who have their complete tenure home in another department
- Partial tenure in the data science initiative with a partner department somewhere else in the university.
- Full tenure home in the data science initiative.

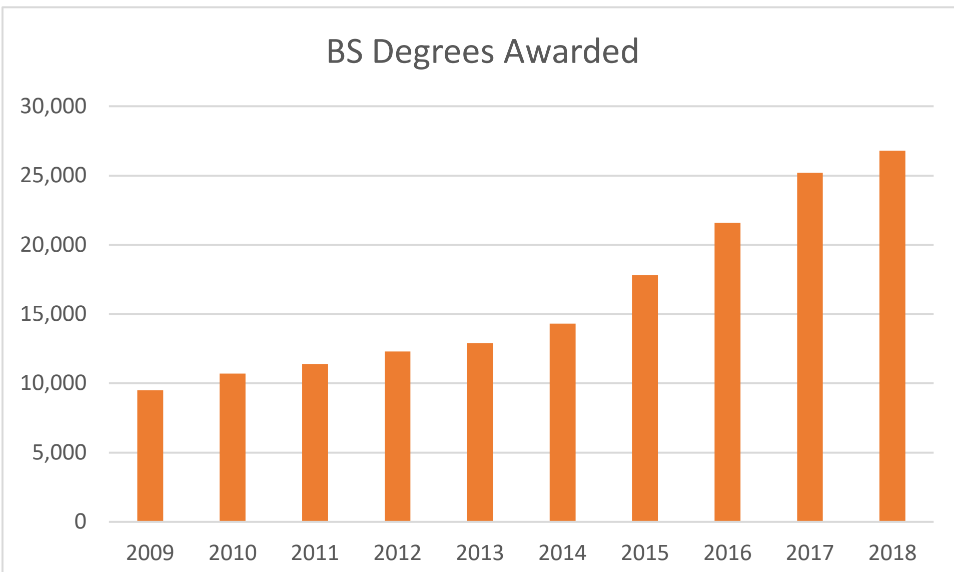
For non-tenure track faculty (NTTF), these affiliations could be the following:

- Affiliation for NTTF of existing or new faculty with appointments in other units in a manner that is similar to the affiliations of TTF (see above).
- Appointments of research teams (e.g. research assistants or postdoctoral scholars) working with a senior TTF or NTTF data science researcher.
- As the appointment home for more senior research scholars such as Research Associates or Research Professors.

# 5 Bringing it all Together

The Computer and Information Science Department is growing in size to address several requirements:

1. Enrollments in Computer and Information Science have been growing year on year for several years. This is a US-wide phenomenon, with the Computing Research Association indicating that "For the tenth consecutive year, there was an increase in the number of new undergraduate computing majors despite the capacity pressures facing departments." The following graph shows this nationwide trend.




2. A critical, core component of data science is the application of machine learning algorithms to domain data. Sometimes this can be done with existing algorithms or tools; more often, successful inference requires that the machine learning algorithms be modified to reflect model aspects of the data. We are particularly focused on the size and breadth of our machine learning faculty.
3. Besides machine learning, high-performance computing and visualization expertise are critical to successful data science endeavors. CIS has strong competence in these areas.

4. In addition to the explosion of data science-related jobs, there is a critical shortage of graduates with sufficient skills in cybersecurity. The department is proposing a curriculum for a specialization in cybersecurity, and looking to grow its faculty in that area in order to deliver that specialization curriculum.

Additionally, the head of department and some of the senior faculty members have been involved, with CAS development, in discussions with existing and potential donors. The recent Ripple gift in cybersecurity is one example of those efforts, as are significant equipment donations by NVIDIA to the NSF Center for Big Learning.

The data science initiative highlights that "computing everywhere" is true for UO. The assertion by the visioning committee that the program not be housed in any existing college or school indicates that an additional college/school-level entity needs to be created. The strong coupling of the initiative with the machine learning and high-performance computing expertise within CIS demonstrates that there is a strong affinity for CIS to be in the same college/school structure as the data science program.

There may be many other individuals, groups, programs, and departments in existing units that have a strong affinity with a School of Computer and Data Sciences. If a green light is given to this proposal, a task force/committee should be charged to determine the identities of those affinite[4] units and the appropriateness for inclusion in the School.

[4] I apologize for creating a new word in the English language. I am amazed that there is not an adjectival form to apply to an entity that has an affinity for another entity.