

Received 2/9/22  
Academic Senate

Final  
To Division  
May 2022

*Proposal for a*  
**MASTER OF SCIENCE DEGREE  
IN  
COMPUTATIONAL DATA SCIENCE**



Marlan and Rosemary Bourns College of Engineering

University of California – Riverside

Riverside, CA 92521

Submitted by

Vassilis Tsotras

(on behalf of the MS in Computational Data Science Program Committee)

Professor, Department of Computer Science and Engineering

Director, Data Science Center

Program Director, Data Science Undergraduate Program

M.S. in Computational Data Science Approvals

<u>Approvals</u>	<u>Dates</u>

Contact Information:

For any questions, please contact:

Vassilis J. Tsotras

Department of Computer Science and Engineering  
Bourns College of Engineering  
University of California Riverside, CA 92521 USA

Phone: 951-827-2888  
Email: [tsotras@cs.ucr.edu](mailto:tsotras@cs.ucr.edu)

## **Table of Contents**

<b>EXECUTIVE SUMMARY</b>	<b>5</b>
<b>SECTION 1: INTRODUCTION</b>	<b>7</b>
<b>SECTION 2: PROGRAM</b>	<b>26</b>
<b>SECTION 3: PROJECTED NEEDS</b>	<b>30</b>
<b>SECTION 4: PROGRAM FACULTY AND STAFF</b>	<b>33</b>
<b>SECTION 5: COURSES</b>	<b>34</b>
<b>SECTION 6: RESOURCE REQUIREMENTS</b>	<b>39</b>
<b>SECTION 7: GRADUATE STUDENT SUPPORT</b>	<b>40</b>
<b>SECTION 8: GOVERNANCE</b>	<b>41</b>
<b>SECTION 9: SENATE REGULATION CHANGES</b>	<b>42</b>
<b>APPENDIX A: PROGRAM BYLAWS</b>	<b>43</b>
<b>APPENDIX B: NEW COURSE SYLLABI</b>	<b>47</b>
<b>APPENDIX C: LETTERS OF SUPPORT</b>	<b>57</b>

## EXECUTIVE SUMMARY

This document is a proposal for a (state-supported) Master of Science (M.S.) degree in **Computational Data Science** (CDS), which will be jointly managed by the departments of Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE). Degree requirements and administration of the program are described in this document.

Data has become ubiquitous in everyday life, impacting every profession, including manufacturing, logistics, health care, public safety, and the military. Data also permeates all aspects of science, engineering, and other academic disciplines. As a result, the field of Data Science has emerged as a new academic discipline: the study of data itself. Data Science deals with obtaining insight and information from the analysis of large collections of data. Data Science lies at the intersection of Computer Science and Statistics, and its solutions already serve a variety of application domains in science, engineering and business. There are two programs currently at UCR that relate to Data Science: (i) a new undergraduate Major in Data Science (a collaboration between the CSE and the Statistics Departments), and (ii) a new MS in Business Analytics (a collaboration between the Business School and the Statistics Department; see <https://business.ucr.edu/msba>). The undergraduate Major in Data Science teaches students the basics from statistics and computer science needed to become a data scientist. The MS in Business Analytics focuses more on applications of statistical analysis to business data.

Instead, the MS in Computational Data Science focuses on the **computational foundations of data science**, providing an in-depth understanding of the algorithms and data structures for storing, manipulating, visualizing and learning from large data sets. The proposed MS in CDS is a comprehensive program studying how data can be collected, transformed, analyzed, and used to solve problems across many application areas.

At UCR, relevant courses related to data management, data mining, information retrieval, big data, machine learning, and artificial intelligence have been offered in the Computer Science & Engineering and the Electrical & Computer Engineering Departments. These courses are regularly offered and are very popular. However, our current MS curricula in BCOE do not permit students to obtain a focused mastery of Computational Data Science.

## Proposal for M.S. Degree in Computational Data Science

The proposed program will allow students with an undergraduate degree from a quantitative field, some experience in algorithms and software engineering, and an exposure to introductory statistics (undergraduate level) to enroll in a masters-level program in the general field of Data Science that will grant them a broad understanding of the computational part of the subject.

The new program will rely on existing faculty and will be built mostly on existing courses (only three new courses will be added) within the two departments. It will leverage upon existing facilities in the two departments. Future course offerings will also be through CSE and ECE and the program faculty will be from these departments.

## SECTION 1: INTRODUCTION

### 1.1 Program Objectives

The objective of the MS in Computational Data Science program is to provide training in various aspects of the data lifecycle. Students will gain exposure to data collection, data cleaning, data integration, data management, and data visualization, as well as the theories and techniques necessary for data analysis from data mining, machine learning, information retrieval, and artificial intelligence.

The program aims to admit students from various backgrounds with undergraduate training in quantitative fields (e.g., engineering, physics, math, statistics). We expect that applicants will have some experience in programming, software engineering, and algorithms, and some exposure in probability/statistics. The committee overseeing the formation of the program has considered this aspect very carefully and designed a program that provides both breadth and depth. Two new courses were designed with this purpose in mind: They introduce students from different backgrounds to the basic tools and theory in the Data Science field. Students will then be exposed to the breadth of the area through a set of core courses. They will also be able to focus on various aspects of computational data science and gain in-depth knowledge through specific electives. At the end, students will complete a capstone project (new course) where they will combine technical, analytic, and interpretive skills to design and execute a large-scale data science project that has a focus on real-world applications.

It is also possible to accept students whose undergraduate education did not include the expected experience in programming, software engineering etc. Examples are students whose undergraduate degrees are in chemistry, biology, economics or sociology. Such students may still be admitted to the program with the stipulation that they complete missing courses at the undergraduate level at UCR. The CSE and ECE departments are working on a sequence of “Bridge” summer courses (taught by UCR faculty) that could be used as a first step by students who need instruction in undergraduate fundamentals, such as programming, algorithms, and data structures, prior to entering graduate programs in CSE, ECE, or Computational Data Science. Through these Bridge summer courses, students without the appropriate background can still finish their MS degree in Computational Data Science within 2 years. We expect the Bridge summer courses will increase the reach of this Data Science related MS program in the near future.

## 1.2 Historical Development of Data Science and Departmental Strengths

We live in a world where data is being generated continuously by scientific experiments, digital processes, sensors, social media, mobile devices, etc. The term “big data” refers to data that is arriving from multiple sources at an alarming volume, velocity, and variety. Data Science is a new field that deals with the management of and extraction of knowledge from big data. As a scientific field, Data Science affects research in many domains, including biological sciences, physical sciences, social sciences, and humanities. The importance of Data Science is evident by various related UC-wide initiatives. As an example, UCB has recently created a separate Data Science Division (<https://data.berkeley.edu/>).

The White House “*Big Data Research and Development Initiative*” committed \$200 million to “extract knowledge and insights from large and complex collections of digital data, accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning.” NIH launched the *Big Data To Knowledge (BD2K)* initiative “to enable biomedical research as a digital research enterprise, to facilitate discovery and support new knowledge.” *Harnessing the Data Revolution* is part of [NSF's 10 Big Ideas](#). In particular, “Engaging NSF's research community in the pursuit of fundamental research in data science and engineering, the development of a cohesive, federated, national-scale approach to research data infrastructure, and the development of a 21st-century data-capable workforce.” Other funding agencies (DARPA, IARPA, etc.) have similar research initiatives.

In addition to research, Data Science heavily influences economics and business. Data has become ubiquitous in everyday life: It impacts every profession, from entry-level office workers to CEOs, from team coaches to general managers, from accountants to CFOs. Businesses now have data available to them at a scale that is historically unprecedented; harnessing this data for insight on what customers want provides them with a competitive advantage. Traditional companies (Ford, Walmart, General Electric, etc.) today pride themselves as being transformed to big-data businesses.

Fueled by the explosion of data, Data Science related jobs have proliferated and the demand for data scientists is extremely high; moreover, this demand is expected to be strong for years to come. A 2016 McKinsey report forecasted a shortfall of roughly 250,000 data scientists by 2024. Data scientists are the no. 1 most promising job in



## Proposal for M.S. Degree in Computational Data Science

America for 2019, according to a report from LinkedIn. Similarly, according to Glassdoor, a recruiting site, Data Scientist has been the best job in the US (2015-2019) with around 113K median base salary. Three-fifths of the data science and analytics jobs are in the finance and insurance, professional services, and information technology sectors, but the manufacturing, health care, and retail sectors also are hiring significant numbers of data scientists. Further, the graduates of the Computational Data Science MS will be equipped with the skills and knowledge needed to master the next generation of massive information system deployments and analyze the data those systems generate. We thus expect that the new program will be in high demand among students and will serve the UCR community well.

As another indication of the interest in Computational Data Science, we have experienced high demand among graduate students for related courses (Data Mining, Machine Learning, AI, Big Data, etc.) For example, around 70-100 students attended “CS235: Data Mining”, in its last few offerings; similarly “CS236: Database Management”, “CS 226: Big Data Management” and “CS229: Machine Learning” have enrollments around 50-60 students. We expect many of these students would prefer a degree more concentrated on these particular topics, particularly one with a coordinated project to provide hands-on experience. Thus, the proposed MS program will better serve many of our current students.

Preparing the workforce in Computational Data Science is also important for the local community. Here in the Inland Empire, for example, the Naval Surface Warfare Center (NSWC) in Corona has launched the Universal Hub for Big Data, a project to collect and share Navy data, which will require a qualified workforce. Our ability to keep high-tech employers like this in the region depends on our ability to supply professionals capable of satisfying their technical needs. NSWC has recently contacted BCOE expressing strong interest in the proposed MS program. Similarly, we have the support from local industry and government agencies who would be interested in hiring our graduates (see letters of support).

The MS in Computational Data Science will be of interest as a career next path, to the UCR students graduating from the recently approved BS in Data Science program (offered by the Department of Computer Science and Engineering and the Department of Statistics). Another Data Science related MS that such students can currently follow is the recently approved MS in Business Analytics (offered by the School of Business and the Department of Statistics). The MS in Business Analytics focuses on the analysis of business applications and their data using a statistical approach (the statistical part of Data Science). The proposed MS in CDS will focus instead on the computational part of

Data Science. We expect that some of these students will continue to pursue a PhD degree in CSE or ECE. Further, a BS+MS will be a possibility to add in the future to the Data Science Major (using the MS in Business Analytics and the proposed MS in Computational Data Science as two options depending on the student's interest).

We further note that faculty from the CSE and Statistics departments were awarded a grant<sup>1</sup> from the NSF Harnessing the Data Revolution Data Science Corps (HDR DSC) program to create Data Science pathways for Inland Empire students with special focus on underrepresented minorities. We have teamed with 4 local community colleges (the Riverside Community College District and the San Bernardino Valley College) as well as CalState San Bernardino. Through course articulations between the colleges involved, such pathways will start from the Inland Empire community colleges and will feed our DS Major and eventually lead to either the MS in Business Analytics or the proposed MS in Computational Data Science (depending on the student's interest). In addition to the local community colleges and CSUSB, we have partnered with local businesses and government to offer possible capstone projects and internships and thus enhance the student experiences in the proposed Data Science pathways.

We thus believe that the MS in Computational Data Science program will play an important role in educating UCR and Inland Empire students, by building their expertise from solid core knowledge, covering the essentials in managing and analyzing data, as well as covering the applications of Data Science in real life problems.

The CSE and ECE Departments have many faculty that perform research related to Computational Data Science. There are strong research groups working on Big Data, Database Management, Data Mining, Artificial Intelligence, Deep Learning, Time Series, Vision and Visualization. The existing strength was instrumental in the creation of the Data Science Center. Section 4 lists the current program faculty (13 CSE and 7 ECE). Related research is published in the top conferences and journals, and is consistently funded by various grants from NSF, Army, Navy, DARPA and other funding agencies. As another indication of quality, the proposed MS in CDS program faculty includes 10 recipients of prestigious research awards (the NSF CAREER Award, the NSF Research Initiation Award, the Air Force Young Investigator Award), and various IEEE and ACM Fellows. PhD graduates from these groups are very much sought after from the industry (including Google, Amazon, LinkedIn, Microsoft, Facebook etc.)

---

<sup>1</sup> PI: Mariam Salloum; title: "DS-PATH: Data Science Career Pathways in the Inland Empire"; 10-1-2021; \$1.5M.

### **1.3 Enrollment Projections**

The new MS in Computational Data Science will help to increase the overall graduate enrollment in BCOE, which is also a college aim. In Fall 2019, the three MS programs offered by CSE, ECE or both had the following enrollment: CSE 143, ECE 51 and Computer Engineering (CEN) 49 MS students. In Fall 2020, the numbers were slightly less (probably due to the pandemic): CSE 112, ECE 41, CEN 45. We believe that the new MS in Computational Data Science will be at least as popular as the ECE and CEN MS programs.

We thus aim to start with 20 students in the first year of the program and reach a steady state of around 60 students within 5 years. This would be achieved without hurting enrollment in the CSE, CEN and ECE MS programs (or the new MS in Robotics offered by CSE, ECE and Mechanical Engineering), since the MS in Computational Data Science offers a different career path than these other MS programs.

Further, we expect that many of these students will stay on for PhDs in CSE or ECE, thus allowing us to select PhD students who have already been at UCR.

### **1.4 Relation to Other Programs in UCR and the UC System**

MS programs related to Data Science typically come under two categories: MS in (variations of) Data Science and MS in Business Analytics. Moreover, Data Science related MS programs are offered from a variety of departments and/or schools. Below we discuss versions of these programs within the UC and how the proposed MS in Computational Data Science differs.

The proposed MS in Computational Data Science will be a state-supported program focused on students that are interested in the on-campus experience. It is thus different from the existing self-supported BCOE MSOL on-line program that offers a Data Science specialization (among others).

UCR has recently approved a self-supported M.S. in Business Analytics (offered by the School of Business and the Statistics department). This degree is different from our proposed program as it focuses on non-technical aspects of data management and analysis while we are looking at the computational aspects of Data Science.

Within the UC system, the majority of MS programs related to Data Science are

self-supporting graduate professional degree programs (SSGPDPs), typically offered either as MS in Business Analytics or as MS in Data Science. The following discussion is based on the most current list of SSGPDP programs available at UCOP (see: [https://www.ucop.edu/institutional-research-academic-planning/files/all\\_UC\\_SSGPDPs.pdf](https://www.ucop.edu/institutional-research-academic-planning/files/all_UC_SSGPDPs.pdf)).

In particular, UC Berkeley has an on-line M.S. of Information and Data Science (MIDS) that is offered through their School of Information. They further provide the “5th Year Master of Information and Data Science” program, open to UC Berkeley undergraduate students as a path to earning a professional master’s degree in one additional calendar year. UC Berkeley also has a (self-supported) MS in Engineering program through their Electrical Engineering and Computer Science Department, that offers a concentration in Data Science and Systems.

UC Davis has a professional Master of Science in Business Analytics offered by their Graduate School of Management.

UC Irvine has a self-supported M.S. in Business Analytics offered by the School of Business and a new Masters in Data Science (also self-supported) program through the Bren School of Information and Computer Science (that houses also the Department of Statistics).

UCLA offers through the Samueli School of Engineering, an on-line Master of Science in Engineering With Certificate of Specialization in Data Science Engineering. It also offers a professional Master of Science in Business Analytics through the Anderson School of Management. There is also a professional MS in Applied Statistics (focused on data science and quantitative analytics) offered by the Department of Statistics.

UC San Diego has a (self-supported) Master of Science in Business Analytics that runs through the Rady School of Management and a Masters of Advanced Studies in Data Science and Engineering offered through their Engineering school (that runs over Fridays/Saturdays).

As for UC state-supported MS programs related to Data Science, we could identify four: (1) UC Berkeley has a 11-month in-person Master of Analytics program that focuses on data-driven analytical methods for optimization, statistics, risk management, offered by the Department of Industrial Engineering and Operations Research. (2) At UC Davis there is a Data Science track within the M.S. in Statistics offered by the Statistics Department. (3) A similar approach is taken at UCSB: there is a Data Science

concentration within the M.A. in Statistics offered by the Department of Statistics and Applied Probability. (4) At UCSD there is a Machine Learning & Data Science emphasis offered within the MS degree in the Electrical and Computer Engineering Department. Finally, we note that UCSC allows for designated emphasis on Data Science to PhD students (which is like a minor for graduate studies); however it is not clear whether this emphasis applies to an MS degree as well.

The proposed MS in Computational Data Science will be a unique offering within the UC system given its focus on the computational side of data science. The closest in nature is UCSD's MS in Machine Learning & Data Science. There are some differences in the focus of the two programs; for example, the proposed UCR program requires more computational core courses and covers some of the more advanced/theoretical concepts in its elective list A. This has been our focus from the start (and is thus reflected in the name of the proposed program).

MS in Data Science degrees are also offered by many top universities around the country that have strong research in this field. Examples include the Master in Computational Data Science at CMU, the Masters in Data Science at NYU and the Masters in Data Science at Columbia University.

### **1.5 Contributions to Diversity**

The lack of diversity in computer science and in the information technology sector of the economy, especially among women and underrepresented minorities, is a well-recognized challenge. While the centrality of computing has manifested itself in dramatic increases in enrollment, computing is among the least diverse disciplines in terms of both gender and minority representation<sup>2</sup>. This has clear effects in the workforce. A recent industry report<sup>3</sup> found that improving ethnic and gender diversity in the U.S. technology workforce represents a massive economic opportunity that could create \$470 – \$570Bn in new value for the tech industry, and could add 1.2 – 1.6% to the national GDP. The report identifies underrepresentation of African American and Latino/Latina workers in the tech industry compared to the U.S. workforce as a whole, accounting for 7 and 8 percent of tech workers, respectively, compared to 12 and 16

---

<sup>2</sup> National Academies of Sciences, Engineering, and Medicine. 2018. Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments. Washington, DC: The National Academies Press.<https://doi.org/10.17226/24926>.

<sup>3</sup> Intel Corporation and Dalberg Global Development Advisors (2016); Decoding Diversity: The Financial And Economic Returns to Diversity in Tech; <https://dalberg.com/wp-content/uploads/2016/06/Diversity-report.pdf>

## Proposal for M.S. Degree in Computational Data Science

percent of all U.S. workers. The gap is even larger for women, who represent only 28 percent of the tech workforce compared to 47 percent of the overall labor force.

The table below depicts the number of females and URMs, among all the MS programs offered by CSE and ECE, as well as the BCOE totals (as of Fall 2020 taken from UCR's Institutional Research website<sup>4</sup>). Even though concentrated in the MS degrees only, we can see similar trends with respect to diversity.

Degree	Enrollment
BCOE (M.S.)	373 (22.8% women, 12.6% URM)
Computer Science (M.S.)	112 (22.3% women, 5% URM)
Computer Engineering (M.S.)	45 (20% women, 6.6% URM)
Electrical Engineering (M.S.)	41 (19.5% women, 17% URM)

National MS enrollment data from the Taulbee survey<sup>5</sup>, show higher % in the enrollment of women (but still below 30%) and lower % for URMs (note data does not include Electrical Engineering).

Degree	Enrollment
Computer Science (M.S.)	26.6% women, 2.8% URM
Computer Engineering (M.S.)	29.9% women, 4.6% URM

Because of its ubiquitousness and inherent interdisciplinarity, Data Science has an enormous, and still largely untapped, potential for increasing diversity in computing. While no data was directly available for MS in Data Science degrees, we expect that the Data Science diversity enrollment will be much better than the traditional computer science and electrical engineering fields.

**Vision:** With the proposed MS in Computational Data Science we will further improve diversity in computing by increasing the participation of women and underrepresented minorities. We first aim at achieving the BCOE enrollments for women (22.8%) and URMs (12.6%) within 2 years from the program's start. Our aim is to achieve women

---

<sup>4</sup> <https://ir.ucr.edu/stats/enroll/demographic>

<sup>5</sup>Taulbee Survey, created by the Computing Research Association, CRA:  
<https://cra.org/wp-content/uploads/2021/05/2020-CRA-Taulbee-Survey.pdf>

enrollment close to 30% by year 4 and raise it to around 40% by year 5. Similarly we aim at 15% URM's by year 4 and reaching 20% of URM enrollment by year 5.

We will follow recent research that has suggested multiple 'best practices' for recruitment, retention, and success of diverse student populations in STEM and computing disciplines. Such best practices include<sup>6,7</sup>:

- (BP1) *Minimize or remove the presence of ambient stereotypes in the classroom* [CP+2009];
- (BP2) *Include real-world context in the curriculum* [BD2016],[DF2007];
- (BP3) *Emphasize collaborative problem solving and interdisciplinary projects* [PH+2007],[PW2009];
- (BP4) *Provide multiple pathways to attract computing majors* [PH+2007];
- (BP5) *Enable student participation in undergraduate research programs* [CS+2014],[CC+2009],[DF2007],[GTH2016],[VG+2013],[WH+2012];
- (BP6) *Provide introductory course options that separate students with prior experience in computer science from those with no experience* [CL2011];
- (BP7) *Provide learning environments other than those expected in the normative culture of computer science* [SA+2015];
- (BP8) *Enlist a diverse team of faculty, instructors, and mentors;*
- (BP9) *Perform active assessment and refinement of institutional culture and practices with an eye to inclusivity.*

**Plan:** To achieve our vision we will follow a multipronged approach that addresses diversity at different levels, in: *recruitment, curriculum and pedagogy, outreach, assessment and faculty*. Below we discuss the various efforts and how they follow the best-practices listed above (BP1-BP9).

**Recruitment:** UCR is an accredited Hispanic Serving Institution (OPEID 00131600), with approximately 35% Hispanic enrollment. The MS in Computational Data Science program aims to recruit from the four closely related UCR BS programs—Data Science, ECE, CEN, and CSE. The fast-growing nature of the field of Data Science (and AI in general) is a great motivating factor for these students to complete a MS degree before entering the workforce. In addition to UCR, the MS in Computational Data Science

---

<sup>6</sup> National Academies of Sciences, Engineering, and Medicine. 2018. Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments. Washington, DC: The National Academies Press. <https://doi.org/10.17226/24926>.

<sup>7</sup> National Academies of Sciences, Engineering, and Medicine. 2020. Roundtable on Data Science Postsecondary Education: A Compilation of Meeting Highlights. Washington, DC: The National Academies Press. <https://www.nap.edu/read/25804>.

program will recruit from other local HSI, such as California State University campuses in the Southern California region (e.g., San Bernardino, Cal Poly Pomona, etc.), California Baptist University, La Sierra University. It is expected that some of the MS students will stay on for a PhD in CSE or ECE, thus enhancing diversity in the associated PhD programs too.

The CSE and ECE departments are heavily invested in broadening participation in computing and are working on increasing diversity at the undergraduate and graduate level through several initiatives. We note that the departments have received various grants for improving diversity. Such efforts will help encourage students (especially traditionally underrepresented minorities and first-generation college students) to consider graduate school as an option. Below we summarize the most recent of these grants and the related activities.

(1) BCOE has received two Google exploreCSR awards (Prof. Mariam Salloum (CSE) in 2020 and Prof. Basak Guler (ECE) in 2021). Google's exploreCSR awards aid higher education efforts to support students from historically marginalized groups to pursue graduate studies and research careers in computing. As part of the exploreCSR program, we are holding several workshops that include: “demystify the graduate school application process” (M.S. and Ph.D.), “explore faculty research” panel, and “graduate student experiences” panel. The program also matches selected students with faculty mentors for a summer research experience. Prof. Mariam Salloum (CSE) piloted the exploreCSR program last year and supported 8 students for a summer research experience (4 female, 1 African American, 2 Hispanics). The activities created as part of the exploreCSR grants align with best practice BP5 (create opportunities for students to participate in undergraduate research programs).

(2) BCOE was also involved in a recently awarded NSF grant<sup>8</sup> to extend the Computing Alliance for Hispanic Serving Institutions (CAHSI). This Broadening Participation in Computing (BPC-AE) grant aims to encourage traditionally underrepresented students (in particular Hispanics and women) to pursue graduate studies. The objectives of the project are: (1) expand students' research knowledge at participating CAHSI institutions through exposure to research as a career pathway; (2) involve upper division students in research experiences that prepare them for graduate-level research; and (3) build graduate program infrastructure through adoption and study of evidence-based, multi-institutional graduate support structures that lead to underrepresented students'

---

<sup>8</sup> BPC-AE: An Extended CAHSI Alliance to Broaden Participation in Graduate Studies”, PI Ann Gates, University of Texas at El Paso; \$2.9M; UCR portion: \$135K.



success in their graduate studies. The above activities align with best practice BP4 (adding career pathways) and BP5 (undergraduate research).

(3) Faculty in CSE were awarded the Northeastern Center of Inclusive Best Practices grant<sup>9</sup> to broaden participation in computing. The project aims to increase the number of female students graduating with computing majors. To achieve this goal, the grant introduced a new introductory course sequence that will lead to Computer Science or Data Science major. The project activities also involve faculty workshops on inclusive pedagogy and equity and data collection. To tackle best practices BP1 and BP7 and change the culture in the classroom, we recognize we must first change the departmental cultures by providing faculty proper training. The faculty workshops are focused on inclusive teaching practices and updating course material to ensure inclusivity (for example, omitting the use of the 'Lena' photo in the image processing field). The workshops will highlight strategies for creating a culture that supports all students, borrowing ideas from *"Twelve Tips for Creating a Culture that Supports All Students in Computing"* by C. Lewis.

(4) A report by the National Academies of Sciences, Engineering, and Medicine [Nat2018] suggested that to build broad participation, Data Science programs should create multiple entry points and allow students with various backgrounds to pursue Data Science, which aligns well with best practices (specifically BP4 and BP6) which outline the importance of creating multiple pathways. Following this suggestion, the Bridge summer program (mentioned earlier, under creation by CSE and ECE) would offer courses so that students with non-CS degrees can cover key undergraduate material and then proceed to pursue an M.S. in Computer Science, Computer Engineering, or Computational Data Science. We expect that this path will be attractive to students from fields like Biology/Neuroscience, Sociology, Economics, Earth Sciences etc. Such entry points will be instrumental in improving diversity. Using gender as an example (we will in fact look at other diversity metrics as well), in 2020 CSE and ECE had only 18% and 12% females respectively, in comparison to 32% in Economics, 75% in Sociology, 66% in Bio/Neuroscience and 54% in Earth Sciences. Recent research has shown that similar Bridge summer coursework, like the one at Northeastern University [BM+2020], increases diversity at the M.S. level. In fact, UCR is part of the M.S. Pathways to Computing Consortium<sup>10</sup> led by Northeastern University to create Bridge programs that lead to graduate level degree programs in Data Science and Computing with the aim of

---

<sup>9</sup> Northeastern CIC Best Practices grant, PI Tamar Shinar, Co-PI Jiasi Chen, Co-PI Paea LePendou, Co-PI Mariam Salloum; \$607K.

<sup>10</sup><https://www.khoury.northeastern.edu/information-for-overview/prospective-align-pillar/ms-pathways-to-computing-consortium/>

Broadening Participation. This program has received initial funding from Google and The Cognizant U.S. Foundation to be awarded in the form of scholarships for students enrolled in M.S. Bridge programs at participating institutions (which includes UCR). These scholarships will be open to students wishing to pursue the Bridge program at UCR and are reserved for traditionally underrepresented minorities.

(5) As mentioned earlier, UCR was recently awarded the NSF Data Science Corps grant<sup>11</sup> to create pathways into undergraduate and graduate Data Science degrees. The program is a partnership between UCR, CSU San Bernardino, and 4 community colleges (Riverside City College, Norco College, Moreno Valley College, and San Bernardino Valley College). Project initiatives include creating pathways between 2-year programs, 4-year programs and graduate programs through course articulation, recruitment and mentorship (BP4). The program has other components, including K-12 outreach, summer fellowship program (BP7), professional development workshops that focus on interview preparation, imposter syndrome, oral-communication, etc. and faculty and teacher workshops.

Curriculum and pedagogy: Prior research has established that the hands-on aspects of Data Science projects, coupled with clear workforce-related applications, appeal to students from underrepresented backgrounds and increase engagement, involvement, and retention [Lue2019, Wil2017, Raw2018]. The cross-disciplinary nature of Data Science creates the opportunity to build a curriculum that is sensitive to the needs of diverse learners as well as diverse members of society. Of particular interest and concern is the subject of algorithmic bias in AI and machine learning. The design of AI systems has been primarily the domain of white, male engineers [6], and several scholars have suggested that efforts toward inclusion in the ranks of those who design AI systems could reduce bias [JIV2019, Wor2020]. For example, just 12% of machine learning engineers are women [Wir2018], with Black AI leaders pointing to a “diversity crisis” in the field [Sno2018]. Critiques of simple inclusivity efforts suggest that diversity programs cannot address overlapping forms of inequality, and have called for applying a more deliberate lens of intersectionality to the algorithm design [Cis2019, DK2020]. To this end, the MS in Computational Data Science program will work closely with the CSE and ECE Departments to update their AI, machine learning, and computer vision graduate-level curricula to address algorithmic bias, including how biased vs. unbiased algorithms may impact society as the technology evolves.

---

<sup>11</sup> NSF Data Science Corps grant “DS-PATH: Creating Data Science Career Pathways,” PI Mariam Salloum, Co-PI Xinping Cui, Co-PI Analisa Flores, Co-PI Vassilis Tsotras, Co-PI Paea LePendou.

Another way to increase participation and retention is to avoid filter or gate-keeping courses (especially early in the program) and replace them with courses that entice student participation through heightening the excitement and applicability of Data Science (BP6). This is the guiding principle behind the design of “CS252A/EE251A: Data Analytics and Exploration” and “CS252B/EE251B: Fundamentals of Data Science”, which jointly serve as an introduction to Data Science through interesting and application-based projects and homework assignments. Moreover, the capstone course “CS279/EE279” leverages Project-Based-Learning (PBL) which allows students to collaborate in teams on real-projects proposed by partnering organizations (industry, local agencies like the County of Riverside or City of Riverside, non-profit organizations, etc.). The real and applied projects integrated into the coursework serve an important component and align well with best practices BP2 and BP3.

We also note that UCR is a member of the Technology Pathway Initiative (TPI) that is led by CAWIT (Center for Advancing Women in Technology; <https://www.cawit.org/>). CAWIT aims to increase the participation of women in computing and information technology, by developing new interdisciplinary computing degree programs that educate more women innovators for the Digital Age. Participating TPI campuses (including UCB, UC Davis and various Cal States) meet at the annual Interdisciplinary Computing Summer Institute (ICSI) to discuss curriculum and pedagogy with respect to diversity, equity and inclusion. We will use input from these meetings to guide continuous improvements to our curriculum based on experiences in other participant campuses.

Recent research [KI+2018] that has stemmed from ICSI has shown that interdisciplinary degree programs are successful in addressing the gender gap in computing. For example, San Francisco State University (also a TPI member) created the Promoting Inclusivity in Computing (PINC) program, which allows students to complete a minor in ‘Computing Applications.’ The program is targeted toward Biology and Chemistry majors, but all non-CS majors are welcome to participate. Based on data reported by the PINC program, 73% of students are women and 51% are underrepresented minorities (URM). While little data exists for M.S. programs, we believe that similar success can be achieved at the graduate level. We will thus reach out to other UCR departments to create interdisciplinary courses that will provide our M.S. students an opportunity to apply their Data Science skills to various domain areas. Such courses can be added in the list of electives offered by the proposed program.

Outreach: Students in the MS in Computational Data Science program will be encouraged to participate with ongoing efforts at UCR to provide mentorship and

broaden participation in Data Science related activities. BCOE student organizations such as the IEEE@UCR and ACM@UCR have a long and successful history of outreach efforts to the local community which bring K-12 students to campus. The MS in Computational Data Science program will work with BCOE student organizations to create new opportunities for graduate student leadership and participation surrounding community outreach events.

For example, graduate students can participate in hackathons as workshop leaders, mentors, or judges. RoseHack is a women-centric hackathon run by Women in Computing (WinC) that involves ~250 students from UCR, surrounding CSU and Community College campuses as well as high schools. Another example are the annual summer robotics camps organized by the ECE department for middle and high schools students, in collaboration with the Redlands Unified School District. Data Science expertise enables the decision making of robots and is an essential component of robots and other intelligent systems. We will encourage students from the MS in Computational Data Science program to take a leadership role in these outreach efforts and mentor the next generation of students.

There is a recent surge of workshops and conferences that promote diversity in Data Science and related fields, with prominent examples including “WiML” (Women in Machine Learning; <https://wimlworkshop.org/>), “WiDS” (Women in Data Science; <https://www.widsconference.org/>), and “BPDM” (Broadening Participation in Data Mining; <https://www.facebook.com/BPDMProgram>). The WiDS conference surpassed 100,000 attendees in hundreds of cities this past March. UCR participates in WiDS as an ambassador and holds UCR-based WiDS-sponsored workshops where our students present and share their work (most recently in March 2021). This will be a great opportunity to create visibility within the larger community that portrays Data Science as a viable career choice for female students. We will also invite undergraduate and high-school students so they might learn about Data Science and hear talks about projects / research from graduate students and faculty.

Continued Assessment of Diversity: The oversight committee of the MS in Computational Data Science program will organize an open feedback session at the end of each academic year in order to obtain qualitative feedback from students and instructors. In addition, the committee will perform quantitative diversity assessment through anonymous student survey and evaluation, in collaboration with the two participating departments. The continued assessment aligns well with best practice BP9 in best practices which outlines that institutions should evaluate enrollment and refine

institutional culture and practices with a focus on broadening participation and inclusivity.

Moreover, the oversight committee will work closely with the CSE and ECE Broadening Participation Committees (BPC) to track and evaluate enrollment and retention data. Assessment of diversity is one of the 5 goals of the CSE and ECE Broadening Participation Committees (BPC)<sup>12</sup>. In particular, the two departments will continue data collection to understand the effect of various interventions and activities of BPC so as to increase the number of women and underrepresented minorities pursuing graduate degrees (goal #4).

Further, assessment of diversity is a core component of the recent NSF Data Science Corps (DS Corps) grant and as part of the grant activities data will be collected with the help of Institutional Research to evaluate admission, demographics and retention. Grant evaluation also includes qualitative assessment through anonymous surveys distributed to students enrolled in the program. The data and report generated by the DS Corps grant will be shared with the oversight committee for review.

Improving Faculty Diversity: It is well known that improving faculty diversity will provide more role models to the program's students and even attract a more diverse student population. Role models have an important position to play, especially for students considering career choices. Improving the diversity of faculty and role models is also highlighted as one of the best practices (BP8) that contribute toward broadening participation. The proposed MS in CDS program faculty includes 5 females and two URM's (1 Pacific Islander and 1 Latino). Even though the proposed program is not a department that can hire its own faculty, there are currently three faculty searches in the related departments (CSE and ECE) in areas related to Data Science. Special efforts will be taken for these and future searches to: (i) increase the pool of candidates with minorities (by announcing the recruitments in related forums including WiML, BlackInAI, LatinXinAI, QueerInAI, HSI institutions and the National Society of Black Engineers) and, (ii) include females/minorities in the search committees in an effort to eliminate unconscious bias.

In particular, BCOE has recognized the importance of further improving diversity among its faculty. Recently, BCOE was awarded an NSF ADVANCE grant<sup>13</sup> to support further recruitment and success of diverse faculty. The NSF ADVANCE program is designed to

---

<sup>12</sup> <https://bit.ly/cse-bpc>

<sup>13</sup> PI: Kelley Barsanti; title: "ADVANCE Partnership: Promoting Equity and Inclusion to Facilitate Retention of Faculty through Evidence- and Place-Based Intervention Training"; 10/1/2021; \$802K.

foster gender equity through a focus on the identification and elimination of organizational barriers that impede the full participation and advancement of diverse faculty in academic institutions. Important elements in this project for supporting systemic and sustainable change include: engagement of leadership at the College, University, and UC System levels; development of organizational leaders within departments as “allies”; and faculty empowerment through department-centralized training (train-the-trainer model) in bystander intervention and unconscious bias. Bystander intervention training is one promising approach for achieving positive changes in individuals and systemic and sustained improvements in organizational climate and culture, to mitigate exclusionary behavior and promote inclusion and retention. Training in unconscious bias is also very important for faculty in search committees.

We will work closely with the BCOE ADVANCE team to improve diversity through faculty recruitment and retention. Faculty in the proposed MS in CDS program will follow the recommendations of the ADVANCE team, including training in unconscious bias and bystander intervention. Further, ADVANCE will develop a group of organizational leaders (“allies”) within the BCOE departments that would model behaviors that promote equity and inclusion. Prof. Tsotras is serving as one of the ADVANCE project’s initial faculty “allies” at BCOE.

#### References:

- [BD2016] Brinkman, B., and A. Diekman. 2016. Applying the Communal Goal Congruity Perspective to Enhance Diversity and Inclusion in Undergraduate Computing Degrees. Paper presented at SIGCSE 2016. Memphis, TN. March 2-5.
- [BM+2020] Brodley Carla, Barry Megan, Aidan Connell, Catherine Gill, Ian Gorton, Benjamin Hescott, Bryan Lackaye, Cynthia LuBien, Leena Razzaq, Amit Shesh, Tiffani Williams, and Andrea Danyluk. 2020. “An MS in CS for non-CS Majors: Moving to Increase Diversity of Thought and Demographics in CS”. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20), pp 1248–1254.
- [CC+2009] Conrad, S., S.S. Canetto, D. Macphee, and S. Farro. 2009. What attracts high-achieving, socio- economically disadvantaged students to the physical sciences and engineering? College Student Journal. 43(4): 1359-1369.
- [Cis2019] Ciston, Sarah (2019-12-29). "Intersectional AI Is Essential". Journal of Science and Technology of the Arts. 11 (2): 3–8. doi:10.7559/citarj.v11i2.665. ISSN 2183-0088.
- [CL2011] Cohoon, J., and T. Luther. 2011. Analysis of a CS1 Approach for Attracting Diverse and Inexperienced Students to Computing Majors. Paper presented at SIGCSE 2011. Dallas, TX. March 9-12.

## Proposal for M.S. Degree in Computational Data Science

- [CP+2009] Cheryan, S., V.C. Plaut, P.G. Davies, and C.M. Steele. 2009. Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*. 97(6): 1045.
- [Cra2016] Crawford, Kate (2016-06-25). "Opinion | Artificial Intelligence's White Guy Problem". *The New York Times*. ISSN 0362-4331.
- [CS+2014] Chang, M.J., J. Sharkness, S. Hurtado, and C.B. Newman. 2014. What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research in Science Teaching*. 51(5): 555-580.
- [DF2007] Davis, C.G., and C.J. Finelli. 2007. Diversity and retention in engineering. *New Directions for Teaching and Learning*. 11: 63-71.
- [DK2020] D'Ignazio, Catherine; Klein, Lauren F. (2020). *Data feminism*. MIT Press. ISBN 0262044005.
- [GTH2016] Gates, A.Q., H. Thiry, and S. Hug. 2016. Reflections: The Computing Alliance of Hispanic-Serving Institutions. *ACM Inroads*. 7(4): 69-73.
- [JIV2019] Jobin, Anna; Ienca, Marcello; Vayena, Effy (2 September 2019). "The global landscape of AI ethics guidelines". *Nature Machine Intelligence*. 1 (9): 389-399. arXiv:1906.11668. doi:10.1038/s42256-019-0088-2.
- [KI+2018] Kulkarni Anagha, Ilmi Yoon, Pleuni S. Pennings, Kazunori Okada, and Carmen Domingo. 2018. Promoting diversity in computing. *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2018)*. Association for Computing Machinery, New York, NY, USA, 236–241.
- [Lue2019] Lue, R. A. (2019). *Data Science as a Foundation for Inclusive Learning*. *Harvard Data Science Review*, 1(2). doi: 10.1162/99608f92.c9267215.
- [Nat2018] National Academies of Sciences, Engineering and Medicine report: "Envisioning the Data Science Discipline: The Undergraduate Perspective", 2018. <https://www.nationalacademies.org/our-work/envisioning-the-data-science-discipline-the-undergraduate-perspective>
- [PH+2007] Peckham, J., L.L. Harlow, D.A. Stuart, B. Silver, H. Mederer, and P.D. Stephenson. 2007. Broadening Participation in Computing: Issues and Challenges. Paper presented at ITiCSE 2007. Dundee, Scotland, UK. June 23-27.
- [PW2009] Pulimood, M., and U. Wolz. 2009. Problem Solving in Community: A Necessary Shift in CS Pedagogy. Paper presented at SIGCSE 2008. Portland, OR. March 12-15.
- [Raw2018] Rawlings-Goss, R. (2018). "Keeping Data Science Broad: Negotiating the Digital and Data Divide Among Higher Education Institutions". *South Big Data Innovation Hub*. [http://bit.ly/ KeepingDataScienceBroad\\_Report](http://bit.ly/KeepingDataScienceBroad_Report).

## Proposal for M.S. Degree in Computational Data Science

[SA+2015] Simon, R.A., M.W. Aulls, H. Dedic, K. Hubbard, and N.C. Hall. 2015. Exploring student persistence in STEM programs: A motivational model. *Canadian Journal of Education*. 38(1): 1-27.

[Sno2018] Snow, Jackie, "'We're in a diversity crisis': cofounder of Black in AI on what's poisoning algorithms in our lives". *MIT Technology Review*. February 14, 2018.

[VG+2013] Villa, E., A. Gates, K. Kephart, S. Hug, and H. Thiry. 2013. Affinity research groups in practice: Apprenticing students in research. *Journal of Engineering Education*. 102(3): 444-466, doi:10.1002/jee.20016.

[WH+2012] Wilson, Z.S., L. Holmes, M.R. Sylvain, L. Batiste, M. Johnson, S.Y. McGuire, S.S. Pang, and I.M. Warner. 2012. Hierarchical mentoring: A transformative strategy for improving diversity and retention in undergraduate STEM disciplines. *Journal of Science Education and Technology*. 21(1): 148-156.

[Wil2017] Williams, T. (2017). "Diversity and Inclusion in Data Science: Using Data-Informed Decisions to Drive Student Success." Webinar Presentation to the Committee on Envisioning the Data Science Discipline: The Undergraduate Perspective, November 7, 2017. <http://www.nas.edu/envisioningDS>.

[Wir2018] Wired: "AI Is the Future—But Where Are the Women?". ISSN 1059-1028. August 17, 2018.

[Wor2020] World Economic Forum: "White Paper: How to Prevent Discriminatory Outcomes in Machine Learning". March 2018. Retrieved December 8, 2020.

### 1.6 Comments from other UC programs

to be updated.

### 1.7 Administration of the Program

The program will be led by a Program Director, assisted by an Associate Director. While the Director will focus on the overall program and coordination among the departments, the Associate Director will serve the role of Graduate Advisor taking care of all graduate student advising issues within the program. A staff member will help the faculty Directors in administering the program. The program faculty will consist of Senate faculty in related research areas from the two departments (see list of the initial program faculty in Section IV). In the interest of efficient administration, a core group of faculty will be appointed to oversee the program and coordinate efforts with the two departments. This Oversight Committee will consist of 5 faculty from the two



## Proposal for M.S. Degree in Computational Data Science

departments (three from CSE and two from ECE), including the Director and Associate Director.

This proposal was created by the following group of faculty:

Samet Oymak (ECE)

Vagelis Papalexakis (CSE)

Mariam Salloum (CSE)

Christian Shelton (CSE)

Amit Roy-Chowdhury (ECE)

Vassilis Tsotras (CSE) - Committee Chair

### **1.8 Evaluation of the Program**

As is the norm for all graduate programs at the UCR campus, the program will follow the Senate-mandated review (once every six or seven years). Beginning with the second year, the Program Committee will initiate an internal review of the M.S. in Computational Data Science Program.

## **SECTION 2: PROGRAM**

Below we describe the undergraduate admission requirements, the program of study and provide a sample time plan.

### **2.1 Admission Requirements**

All applicants to this program must have completed a Bachelor's degree or its approved equivalent from an accredited institution and to have attained undergraduate record that satisfies the standards established by the Graduate Division and University Graduate Council. Students need experience in a quantitative field with experience in programming, software engineering, algorithms, and background in statistics. Competence in these areas is defined by the following UCR undergraduate courses (or equivalents):

- CS 141 - Intermediate Data Structures and Algorithms
- CS 100 - Software Construction
- MATH 010A - Multivariable Calculus
- MATH 031 - Linear Algebra
- A course covering foundations of probability and statistics (such as STAT 155 - Probability and Statistics for Science and Engineering, or, EE 114 - Probability, Random Variables, and Random Processes in Electrical Engineering)

Applicants who fail to meet this criterion may sometimes be admitted with course deficiencies, provided they take remedial steps to cover the deficiencies. A student who is deficient in a competency area may be asked to complete the corresponding UCR course with a letter grade of at least B, or to pass a challenge examination based on that course's final exam with a grade of at least B. All such remedial work cannot be counted towards the MS degree requirements and should be completed within the first year of graduate study, and in all cases the deficiency(s) must be corrected BEFORE a student can enroll in any graduate course from the same specialty area. The details will be decided by the Graduate Advisor of the program in consultation with the student. The CSE and ECE departments are working currently on a sequence of 'Bridge' summer courses that can be used as a first step by students who lack basic undergraduate background in programming, algorithms and data structures.

All applicants must submit scores from the Graduate Record Exam, General Test (GRE). Relevant GRE subject tests may be beneficial to the candidate's application, but are not required. Applicants whose first language is not English are required to submit acceptable scores from the TEST of English as a Foreign Language (TOEFL) or the International English Language Testing System (IELTS) unless they have a degree from an institution where English is the exclusive language of instruction. Additionally each applicant must submit letters of recommendation, as per the admission requirements. All other application requirements are specified in the graduate application.

## **2.2 Computational Data Science MS Program**

The MS in Computational Data Science program will be built using existing courses, and three recently approved courses<sup>14</sup> (CS252A/EE251A: Data Analytics and Exploration, CS252B/EE 251B: Fundamentals of Data Science and CS/EE 279: Capstone Project in Data Science). The MS in Computational Data Science requires the completion of 49 units of coursework, including a capstone project. There are no thesis or comprehensive exam options; i.e. it falls in the category of a Master's II (with capstone).

Units are divided among core courses (6 courses, for a total of 24 units), elective courses (5 courses, for a total of 20 units), a professional development course (1 unit) and the capstone course (4 units). All students must complete the same core courses. Elective courses are selected by the student from a list of possible courses; students can petition to select a course not on the list.

### **Core courses:**

1. CS 252A/EE 251A: Data Analytics and Exploration (recently approved course)
2. CS 252B/EE 251B: Fundamentals of Data Science (recently approved course)
3. CS 224: Fundamentals of Machine Learning
4. CS 226: Big Data Management
5. CS 235: Data Mining Techniques
6. CS 236: Database Management

### **Elective courses:**

The five electives can be selected from the following two lists of elective courses; at least three of the courses must be from list A. The description of all the elective courses is available later in the proposal. Students may petition for other elective courses; such

---

<sup>14</sup> The full description of the three newly approved courses appears in Appendix B.

petitions require approval of the program graduate advisor. We expect that the list of electives will be enhanced in the future by approved UCR courses related to Computational Data Science.

**Elective List A:**

1. CS 205: Artificial Intelligence
2. CS 225: Spatial Computing
3. CS 227: Probabilistic Models for Artificial Intelligence
4. CS/EE 228: Introduction to Deep Learning
5. CS 229: Machine Learning
6. CS 242: Information Retrieval and Web Search
7. CS/EE 248: Optimization for Machine Learning
8. EE 231: Convex Optimization in Engineering Applications
9. EE 236: State and Parameter Estimation Theory
10. EE 240: Pattern Recognition
11. EE 244: Computational Learning

**Elective List B:**

1. CS 210. Scientific Computing
2. CS 211. High Performance Computing
3. CS/EE 217: GPU Architecture and Parallel Programming
4. CS 234: Computational Methods for Biomolecular Data
5. EE 241: Advanced Digital Image Processing
6. EE 243: Computer Vision
7. EE 250: Information Theory

**Capstone Experience:** Students must complete a capstone course CS/EE 279: Capstone Project in Data Science (new course), under the guidance of the capstone instructor member. The description of the capstone course appears in Section 5.

**Professional Development Requirement:** Students will satisfy the professional development requirement by enrolling in one of the following courses: one quarter of CS 287 (Colloquium in Computer Science), or GDIV 403 (Research and Scholarship Ethics), or at least one unit of CS 298I (Individual Internship).

## **2.3 Other Requirements**

There are no field or qualifying examinations. There is no thesis/dissertation or final examination. There are no special requirements over and above the Graduate Division minimum requirements.

## 2.4 Sample Program

Below we provide a sample program. Assuming that a student has no deficiencies and is full-time, the normative time from matriculation to degree is 4 quarters. Using the currently planned Bridge summer courses in CSE/ECE, it is expected that students with deficiencies can still graduate within 2 years.

	Fall	Winter	Spring
Year 1	CS252A/EE 251A CS252B/EE 251B CS 224	CS 226 CS 236 Elective	CS235 Elective Elective
Year 2	Elective Elective CS/EE 279		

## **SECTION 3: PROJECTED NEEDS**

### **3.1 Student Demand and Opportunities**

We expect a large demand for the new program. The numbers of students in related programs, like the BS and MS programs in CSE, ECE and CEN continue to increase. The proportion of domestic students in related MS programs is around 17% in CSE, 20% in CEN and 33% in ECE; we expect that for the MS in CDS the percentage of domestics will be closer to the ECE example. This is because Data Science seems to be popular with domestic students. Moreover, we expect that the proposed program (as well as the MS in Business Analytics) will draw from students who complete the new Data Science B.S. program. We note that the B.S. in DS inaugural class of Fall 2020 and the Fall 2021 incoming freshman/transfer class are almost 100% domestic. Many students in the various Engineering Undergraduate Professional Societies have also expressed interest in Data Science related graduate programs. While we have most of the courses, the structures of the existing programs do not allow them to take the proper set of courses required for specialized training in Computational Data Science. This demand is directly related to opportunities for students after graduation.

### **3.2 Opportunities for Placement of Graduates**

Based on our experience from graduate students working in relevant areas (Databases, Data Mining, Artificial Intelligence, Machine Learning etc.) there is currently very high demand from industry. Moreover, as mentioned in the Introduction, according to Glassdoor, a recruiting site, Data Scientist has been the best job in the US (2015-2019).

### **3.3 Importance to the Discipline**

As a scientific field, Data Science affects research in many domains, including biological sciences, physical sciences, social sciences, and humanities. In addition to the newly approved Data Science undergraduate program, the proposed MS in Computational Data Science allows students to concentrate further on this important field.

### **3.4 Meeting the needs of Society**

Data is an important societal asset. By training more students in Computational Data Science we also create more “citizen scientists”. According to CitizenScience.gov (the official government website dedicated to Citizen Science), a citizen scientist “...participates voluntarily in the scientific process, addressing real-world problems in

ways that may include formulating research questions, conducting scientific experiments, collecting and analyzing data, interpreting results, making new discoveries, developing technologies and applications, and solving complex problems". Such involvement can engage the American public in addressing societal needs and accelerating science, technology, and innovation.

### **3.5 Relation to Research and Faculty Interests**

A critical mass of our faculty are engaged in research and teaching across the full range of areas relevant to the proposed MS program. This is also evident from the fact that almost all the courses for the program already exist at UCR. These areas are already of high interest to faculty. Moreover, faculty is well funded in these research areas.

### **3.6 Program Differentiation**

The MS in Computational Data Science will be a state-supported program. It is thus different from UCR's MSOL program in Data Science as well as other similar online/professional programs in other UC campuses. In addition to being state-supported our program differs due to its concentration in the computational part of Data Science. It is also different from UCR's professional MS in Business Analytics (and other UC similar programs) that focuses on non-technical aspects of data management and analysis while we are looking at the computational side of data analysis.

Among state supported programs, the proposed MS in Computational Data Science is different from the UC Berkeley Master of Analytics program (offered by the Department of Industrial Engineering and Operations Research) since that program focuses on data-driven analytical methods for optimization, statistics, and risk management. It is also different from the Data Science concentrations offered by the masters in Statistics (UC Davis and UC Santa Barbara) given their focus on statistical analysis approaches. The closest state supported program is UCSD's Machine Learning & Data Science. The proposed UCR MS in CDS program requires more computational core courses and covers some of the more advanced/theoretical concepts in its elective list A. This focus is reflected in the name of the proposed program. A further discussion of the relevant UC programs appears in Section 1.4.

In summary, UCR's MS in Computational Data Science will be the first UC MS program under this name, with this important focus.

There are also Data Science related offerings from private southern California Institutions. Examples are: (1) the MS in Information Systems & Technology with concentration in Data Science, offered by Claremont Graduate University, and (2) the Master of Science in Computer Science (Data Science) offered by USC. We believe that we offer a very competitive program from a public institution that concentrates on the computational part of Data Science.



## SECTION 4: PROGRAM FACULTY AND STAFF

The list of the Program Faculty (with a link to their publications) appears below:

### **CSE**

Jiasi Chen (Associate Professor; PhD; <https://dblp.org/pid/35/9005.html>)  
Ahmed Eldawy (Assistant Professor; PhD; <https://dblp.uni-trier.de/pers/hd/e/Eldawy:Ahmed>)  
Vagelis Hristidis (Professor; PhD; <https://dblp.uni-trier.de/pers/hd/h/Hristidis:Vagelis>)  
Eamonn Keogh (Professor; PhD; [https://dblp.uni-trier.de/pers/hd/k/Keogh:Eamonn\\_J](https://dblp.uni-trier.de/pers/hd/k/Keogh:Eamonn_J))  
Paea LePendou (Assistant Teaching Professor; PhD; <https://dblp.uni-trier.de/pers/hd/l/LePendou:Paea>)  
Amr Magdy (Assistant Professor; PhD; [https://dblp.uni-trier.de/pers/m/Magdy\\_0001:Amr.html](https://dblp.uni-trier.de/pers/m/Magdy_0001:Amr.html))  
Evangelos Papalexakis (Associate Professor; PhD;  
[https://dblp.uni-trier.de/pers/hd/p/Papalexakis:Evangelos\\_E](https://dblp.uni-trier.de/pers/hd/p/Papalexakis:Evangelos_E))  
C.V. Ravishankar (Professor; PhD; [https://dblp.uni-trier.de/pers/hd/r/Ravishankar:Chinya\\_V](https://dblp.uni-trier.de/pers/hd/r/Ravishankar:Chinya_V))  
Elaheh Sadredini (Assistant Professor; PhD; <https://dblp.org/search?q=Elaheh+Sadredini>)  
Mariam Salloum (Assistant Teaching Professor; PhD; <https://dblp.uni-trier.de/pers/hd/s/Salloum:Mariam>)  
Christian Shelton (Professor; PhD; [https://dblp.uni-trier.de/pers/hd/s/Shelton:Christian\\_R](https://dblp.uni-trier.de/pers/hd/s/Shelton:Christian_R))  
Vassilis Tsotras (Professor; PhD; <https://dblp.uni-trier.de/search?q=tsotras>)  
Neftali Watkinson (Assistant Teaching Professor, PhD; <https://dblp.uni-trier.de/pid/210/5993.html>)

### **ECE**

Salman Asif (Assistant Professor; PhD; [https://dblp.uni-trier.de/pers/hd/a/Asif:Muhammad\\_Salman](https://dblp.uni-trier.de/pers/hd/a/Asif:Muhammad_Salman))  
Bir Bhanu (Professor; PhD; <https://dblp.uni-trier.de/pers/hd/b/Bhanu:Bir>)  
Jia Chen (Assistant Teaching Professor; PhD;  
<https://scholar.google.com/citations?user=iD98H2sAAAAJ&hl=en>)  
Basak Guler (Assistant Professor; PhD; <https://dblp.org/search?q=basak+guler>)  
Samet Oymak (Assistant Professor; PhD; <https://dblp.uni-trier.de/pers/hd/o/Oymak:Samet>)  
Amit Roy-Chowdhury (Professor; PhD; [https://dblp.uni-trier.de/pers/hd/r/Roy=Chowdhury:Amit\\_K](https://dblp.uni-trier.de/pers/hd/r/Roy=Chowdhury:Amit_K))  
Nanpeng Yu (Assistant Professor; PhD; <https://dblp.uni-trier.de/pers/hd/y/Yu:Nanpeng>)

### **STAFF**

One FTE for administrative support, primarily for graduate student admissions, enrollment and advising. Initial support may be less than 1 FTE, ramping up as the program matures.

### **TEACHING RESOURCES**

The new program is based on existing courses from CSE and ECE. The three new courses are cross-listed between the two departments which will share responsibilities in teaching them. In the Appendix we include letters of support from the two department chairs that also discuss the sharing of the teaching.

## SECTION 5: COURSES

### Core Courses

#### New Courses Developed for the proposed Program (already approved)

**CS 252A /EE 251A.** *Data Analytics and Exploration* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): CS141, CS100; Stat 155 or EE114 or equivalent; graduate standing; or consent of instructor. Covers important algorithms relevant to the lifetime of data from data collection and cleaning to integration, data mining, and analytics. Topics include: sketch algorithms for computing statistics on data streams; mining social graphs including community detection and graph partitioning; Data Science life cycle: techniques on data cleaning, data integration, Exploratory Data Analysis, and visualization.

**CS 252B /EE 251B.** *Fundamentals of Data Science* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): Math 010A, Math 031 or EE 020B, CS100; Stat 155 or EE114; graduate standing; or consent of instructor. Explores theoretical tools in data science and their applications in data data science. Introduces and motivates statistical and computational viewpoints on data analysis. Topics include the manipulation of data as vectors, drawing inferences from data as distributions, and quantifying data uncertainty for data analysis. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS/EE 279.** *Capstone Project in Data Science* (4) Lecture, 1 hour; outside research, extra readings, 9 hours. Prerequisite(s): Enrollment in Master in Data Science. Co-requisites: CS 252A/EE 251A, CS 252B/EE 251B, CS224, CS226, CS235, CS236. Covers combining technical, analytic, and interpretive skills to design and execute a large-scale data science capstone project that has a focus on real-world applications. Provides an opportunity to integrate all of the core skills and concepts learned throughout the program and prepares students for long-term professional success in the field. Emphasizes collaboration and communication in both written and oral form.

### Existing Core Courses

**CS 224:** *Fundamentals of Machine Learning.* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): CS 100, STAT 155, MATH 31; graduate standing. A study of generative and discriminative approaches to machine learning. Topics include probabilistic model fitting, gradient-based loss optimization, regularization, hyper-parameters, and generalization. Includes experience with data science

programming environments, data from practice, and performance metrics. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 226.** *Big-Data Management* (4) Lecture, 3 hours; term paper, 3 hours. Prerequisite(s): CS 166 or CS236; graduate standing. Introduction to the architecture and design of big data management systems. Covers the design of distributed file systems and high throughput databases. Description of popular programming paradigms for big data including MapReduce and Resilient Distributed Datasets. Includes a course project with hands-on experience on open-source big data systems. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 235.** *Data Mining Techniques* (4) Lecture, 3 hours; term paper, 1.5 hours; project, 1.5 hours per week. Prerequisite(s): CS 141, CS 170 is recommended; graduate standing. CS 235 online section; enrollment in the online Master of Science in Engineering program. Provides students with a broad background in the design and use of data mining algorithms and tools. Includes clustering, classification, association rules mining, time series clustering, and Web mining. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 236.** *Database Management Systems* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): CS 141; CS 166 is recommended; graduate standing; or consent of instructor. Covers principles of file systems; architecture of database management systems; data models; and relational databases. Also examines logical and physical design of databases; hardware and software implementation of database systems; and distributed databases (e.g., query processing, concurrences, recovery). May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

### **Electives**

**CS 205.** *Artificial Intelligence* (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 170 or equivalent; graduate standing. Examines knowledge representation and automated reasoning and their use in capturing common sense and expert knowledge. Also addresses predicate and nonmonotonic logics; resolution and term rewriting; reasoning under uncertainty; theorem provers; planning systems; and belief networks. Includes special topics in natural language processing, perception, logic programming, expert systems, and deductive databases. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 210.** *Scientific Computing* (4) Lecture, 4 hours. Prerequisite(s): CS 010B, MATH 010A; MATH 031 or equivalent; graduate standing; or consent of instructor. Utilizes scientific computing in a specific computer science research area. Provides a foundation for pursuit of further studies of special topics in scientific computing. May be

taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 211.** *High Performance Computing* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): CS 161 or consent of instructor. Introduces performance optimization for sequential computer programs. Covers high performance computing on multicore shared memory computers and on distributed memory computing clusters. Also covers high performance scientific libraries and computing application development using pthreads, OpenMP, and Message Passing Interface (MPI) parallel file systems. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS/EE 217.** *GPU Architecture and Parallel Programming* (4) Lecture, 3 hours; consultation, 1 hour. Prerequisite(s): CS 160 with a grade of "C-" or better or consent of instructor. Introduces the popular CUDA based parallel programming environments based on Nvidia GPUs. Covers the basic CUDA memory/threading models. Also covers the common data-parallel programming patterns needed to develop high-performance parallel computing applications. Examines computational thinking; a broader range of parallel execution models; and parallel programming principles. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 225.** *Spatial Computing* (4) Lecture, 3 hours; individualized study, 3 hours. Prerequisite(s): graduate standing; or consent of instructor. Introduction to the spatial computing technologies and techniques. Covers the fundamentals, the present, and the emerging use cases of spatial data analysis systems. Topics include spatial data modelling, spatial relationships, storage, indexing, query processing, and recent trends in the field. Includes a research-oriented project and hands-on experience on spatial technologies. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 227.** *Probabilistic Models for Artificial Intelligence* (4) Lecture, 3 hours; written work, 3 hours. Prerequisite(s): CS 224; graduate standing. Covers methods for representing and reasoning about probability distributions in complex domains. Focuses on graphical models and their extensions such as Bayesian networks, Markov networks, hidden Markov models, and dynamic Bayesian networks. Topics include algorithms for probabilistic inference, learning models from data, and decision making. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS/EE 228:** *Intro to Deep Learning*. (4) Lecture, 3 hours; written work, 3 hours. Prerequisite(s): CS 224 or EE 231 or EE 236 or EE 244 or CS 171 or EE 142; graduate standing; or consent of the instructor. Explores fundamentals of deep neural networks and their applications in various machine learning tasks. Includes the fundamentals of perception, approximation, neural network architectures, loss functions, and generalization. Addresses optimization methods including backpropagation, automatic differentiation, and regularization. Covers non-standard problems including auto-encoders and probabilistic models. Presents applications in machine learning/computer vision. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and

graduate advisor.

**CS 229.** *Machine Learning* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): CS 171 or EE 142 or CS 224; graduate standing. For the CS 229 online section; enrollment in the Online Master-in-Science in Engineering program; graduate standing. A study of supervised machine learning that emphasizes discriminative methods. Covers the areas of regression and classification. Topics include linear methods, instance-based learning, neural networks, kernel machines, and additive models. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor. Credit is awarded for one of the following CS 229 or EE 240.

**CS 234.** *Computational Methods For Biomolecular Data* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): CS 111; CS 141 or CS 218; STAT 155 or STAT 160A; graduate standing. A study of computational and statistical methods aimed at automatically analyzing, clustering, and classifying biomolecular data. Includes combinatorial algorithms for pattern discovery; hidden Markov models for sequence analysis; analysis of expression data; and prediction of the three-dimensional structure of RNA and proteins. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor. Credit is awarded for one of the following CS 234 or CS 144.

**CS 242.** *Information Retrieval and Web Search* (4) Lecture, 3 hours; term paper, 1.5 hours; project, 1.5 hours per week. Prerequisite(s): CS 141, CS 166. CS 242 online section; enrollment in the online Master of Science in Engineering program. Introduces Information Retrieval (IR) principles and techniques for indexing and searching document collections with special emphasis on Web search. Includes text processing, ranking algorithms, search in social networks, search evaluation, and search engines scalability. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS/EE 248:** *Optimization for Machine Learning.* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): CS 229 or EE 231 or EE 244; graduate standing; or consent of the instructor. Explores efficient optimization algorithms for machine learning. Emphasizes fundamental principles, provable guarantees, and contemporary results. Includes fundamentals of optimization (first-order methods, stochastic algorithms, accelerated schemes, non-convex optimization, regularization, and black-box optimization). Also covers connections to statistical learning (empirical risk minimization, finite-sample guarantees, and high-dimensional problems).

**EE 231.** *Convex Optimization in Engineering Applications* (4) Lecture, 3 hours; term paper, 3 hours. Prerequisite(s): EE 230. Covers recognizing and solving convex optimization problems that arise in engineering applications. Explores convex sets, functions, and optimization problems. Includes basics of convex analysis, least-squares, linear and quadratic programs, semidefinite programming, minimax, and other problems. Addresses optimality conditions, duality theory, theorems of alternative and applications, interior-point methods, and applications in engineering.

**EE 236.** *State and Parameter Estimation Theory* (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): EE 215 with a grade of C or better; graduate standing. Covers Fisher information, Cramer-Rao lower bound, efficiency, and sufficient statistics. Addresses minimum variance unbiased, best linear unbiased, maximum likelihood, least squares, maximum a posteriori, and mean-squared estimation. Also covers Weiner and Kalman filtering as well as applications in navigation, signal processing, machine learning, and dynamical systems. Cross-listed with ME 236.

**EE 240.** *Pattern Recognition* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): EE 141 or consent of instructor. EE 240 online section; enrollment in the Online Master-in-Science in Engineering program; graduate standing. Covers basics of pattern recognition techniques. Topics include hypothesis testing, parametric classifiers, parameter estimation, nonparametric density estimation, nonparametric classifiers, feature selection, discriminant analysis, and clustering. Credit is awarded for one of the following EE 240 or CS 229.

**EE 241.** *Advanced Digital Image Processing* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): EE 152 or consent of instructor. Covers advanced topics in digital image processing. Examines image sampling and quantization, image transforms, stochastic image models, image filtering and restoration, and image data compression.

**EE 243.** *Advanced Computer Vision* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): EE 146 or EE 152 or EE 215 or EE 228 or CS 228 or EE 241 or CS 225; graduate standing or consent of instructor. For the EE 243 online section: enrollment in the Online Master-of-Science in Engineering program; graduate standing; or consent of instructor. Study of advanced computer vision including classical- and learning-based approaches. Topics include feature extraction, segmentation, motion analysis and tracking, object and activity recognition, projective geometry, modeling and calibrating cameras, and three-dimensional reconstruction. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**EE 244.** *Computational Learning* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): graduate standing or consent of instructor. Explores fundamental computational learning techniques. Topics include elements of learning systems, inductive learning, analytic learning, case-based learning, genetic learning, connectionist learning, reinforcement learning and integrated learning techniques, and comparison of learning paradigms and applications.

**EE 250.** *Information Theory* (4) Lecture, 3 hours; extra reading, 3 hours. Prerequisite(s): EE 215. An overview of fundamental limitations imposed on communication systems. Topics include Shannon's information measures, weak and strong typicality, lossless data compression, source and channel models and Shannon's coding theorems, channel capacity and the rate-distortion function, Gaussian sources and channels, and limits of communication between multiple terminals.

## **SECTION 6: RESOURCE REQUIREMENTS**

All the technical resources required by the M.S. in Computational Data Science program are already available in and for the two participating departments including computing facilities, library resources, teaching laboratories and research facilities. The only additional resources would be office space and one FTE for administrative support (initial support may be less than 1 FTE, ramping up as the program matures). See letter of support from the BCOE Dean.

## **SECTION 7: GRADUATE STUDENT SUPPORT**

MS in Computational Data Science students are expected to be self-supported. However, GSR and Teaching Assistantships may be available on a case-by-case basis.



## **SECTION 8: GOVERNANCE**

The Program Faculty will consist of Senate faculty in related research areas to Computational Data Science, drawn from the two departments. Program Faculty members shall support the program through instruction of courses, supervision of students, activity in Computational Data Science research, or program administration. All Program Faculty are eligible to vote on matters related to the MS in Computational Data Science Program. All changes to the MS in Computational Data Science Program or curriculum must be approved by a majority of the Program Faculty.

The program will be led by a Program Director, assisted by an Associate Director. The Director is appointed by the Dean of BCOE with consultation from the Program Faculty. The Program Director will rotate among the 2 departments. While the Director will focus on the overall program and coordination among the departments, the Associate Director will serve the role of Graduate Advisor taking care of all graduate student advising issues within the program. A staff will help the faculty Directors in administering the program. The staff will report to the Director and the Director will report to the Dean of BCoE.

A core group of the program faculty (including the Director and Associate Director) will be appointed to form the Oversight Committee, whose task is to oversee the program and coordinate efforts with the departments. The committee will consist of three faculty from CSE and two faculty from ECE.

## **SECTION 9: SENATE REGULATION CHANGES**

No changes in Senate Regulations at the Divisional level or in the Assembly of the Academic Senate will be required.

## **APPENDIX A: PROGRAM BYLAWS**

## **MS in Computational Data Science Program Bylaws**

Creation Date: June 2, 2020

Approval Date:

### **I. Objective**

- A. The MS in Computational Data Science is housed in the Bourns College of Engineering (BCOE), and is a joint program between the departments of Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE).
- B. The objective of the MS in Computational Data Science is to provide training in various aspects of Computational Data Science. Students graduating from the program will gain exposure to the foundational principles underlying the full data lifecycle, from storage to management to analysis.

### **II. Membership**

- A. The faculty associated with the program, called the Program Faculty, is drawn from UCR Senate faculty in related research areas from the CSE and ECE departments.
- B. Program Faculty members shall support the program through instruction of courses, supervision of students, activity in Data Science research, or program administration.
- C. All Program Faculty are eligible to vote on matters related to the MS in Computational Data Science Program.
- D. All changes to the MS in Computational Data Science Program or curriculum must be approved by a majority of the Program Faculty.
- E. UCR Senate faculty outside of CSE and ECE whose research or teaching activities align with the mission of the MS in Computational Data Science are eligible to be Cooperating Faculty in the program. Cooperating Faculty do not have a vote in the program, but are eligible to participate in meetings of the Program Faculty.
- F. Membership Changes
  - 1. Nominations of prospective members to the Program Faculty or Cooperating Faculty may be made by any faculty member in CSE or ECE.
  - 2. New Program Faculty or Cooperating Faculty shall be appointed by a majority vote of the Program Faculty, based on a review of the nomination and the recommendation of the Oversight Committee, defined in III.A below.

3. Members of the Program Faculty may terminate their association with the MS in Computational Data Science Program after so informing the Program Director in writing.
4. Participation as Program Faculty or Cooperating Faculty shall be reviewed every three years to ensure that all members are meeting their obligations to the MS in Computational Data Science Program.

### III. Administration

A. A core group of faculty, called the MS in Computational Data Science Program Oversight Committee, shall oversee the program and coordinate efforts with the departments.

#### B. Composition

1. The Program Oversight Committee is chaired by the Director, or by the Associate Director in the Director's absence.
2. The Program Oversight Committee consists of five (5) members (including the Director and Associate Director), all of whom are members of the Program Faculty.
3. Three (3) faculty from CSE and two (2) faculty from ECE departments shall be on the Oversight Committee. Faculty with joint appointments in multiple departments shall specify the one department they represent.

#### C. Duties

1. The duties of the Director include
  - a. providing overall academic and administrative leadership for the program,
  - b. overseeing the development and implementation of program policies,
  - c. representing the interests of the program to the College, the Campus and University administrators,
  - d. calling and chairing meetings of the program,
  - e. managing the program's budgets,
  - f. ensuring the accuracy of publications related to the program including web pages and catalog copy, and
  - g. coordinating the program's teaching needs with the teaching assignments of the constituent departments.
2. The duties of the Associate director include
  - a. serving as the Graduate Advisor for the MS in Computational Data Science program,
  - b. coordinating administration with the Office of Graduate Studies,
  - c. submitting course change or approval forms, and

d. assisting the Director as needed.

D. Appointments

1. The Dean of BCOE appoints the Director with consultation from the Program Faculty, in a manner consistent with the appointment of other program directors and department chairs. The Director reports to the BCOE Dean.
2. It is expected that Directors should alternate between the two departments. Any exception will require a majority vote of the Oversight Committee.
3. Director appointments are for three (3) years, except when circumstances require otherwise.
4. Members of the Oversight Committee, other than the Director, are nominated and elected by the Program Faculty, in accordance with the provisions of bylaw III.B above.
5. The Associate Director will be appointed by the Director from the membership of the Oversight Committee.

IV. Meetings

A. The Program Faculty

1. The Program Faculty will meet as necessary, but at least once a year.
2. Three or more faculty from the Program Faculty can call a meeting.

B. The Program Oversight Committee

1. The Program Oversight Committee will meet at least once per academic term, on a schedule set by the Director.
2. Three or more faculty from the Program Oversight Committee can call a meeting.

C. Members will be notified of meetings at least a week in advance.

D. A quorum for meetings of the Program Faculty consist of 50% of the Program Faculty.

E. A quorum for meetings of the Program Oversight Committee consist of 4 members of the Program Oversight Committee.

## **APPENDIX B: NEW COURSE SYLLABI**

## CS 252A/EE 251A : Data Analytics and Exploration Spring 2021

**Instructor:** Mariam Salloum / Vagelis Papalexakis

**Contact Info:** [msalloum@cs.ucr.edu](mailto:msalloum@cs.ucr.edu) / [epapalex@cs.ucr.edu](mailto:epapalex@cs.ucr.edu)

### **Credits / Type**

4.0 Units

Lecture: 3 hours

Research (outside): 3 hours

### **Description:**

This course covers important algorithms relevant to the lifetime of data from data collection and cleaning to integration, data mining and analytics. Topics include: sketch algorithms for computing statistics on data streams; mining social graphs, including community detection and graph partitioning; Data Science lifecycle and techniques on data cleaning, data integration, Exploratory Data Analysis, and visualization.

**Prerequisite(s):** CS141, CS100, Stat 155 or EE114 or equivalent.

### **Relevant Textbooks**

- (abbreviated MMD) [Mining of Massive Datasets](#) by Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman
- (abbreviated EDA) [Experimental Design and Analysis](#) by Howard J. Seltman. 2018
- Selected papers (See assigned readings in the schedule)

### **Grading:**

- Homework (x5) 35% (assignments include both a written and programming component)
- Midterm (x2) 40%
- Final Project 25%

### **Tentative Schedule**

Week	Lecture Topics	Readings (Book/Papers)
1	Probability Review. Distributions (single and multi-dimensional), Central-limit Theorem, Expectation, Mean, Variance, Moments,	<a href="https://www.stat.cmu.edu/~hseltman/309/Book/chapter3.pdf">https://www.stat.cmu.edu/~hseltman/309/Book/chapter3.pdf</a> (EDA book Ch.3)
2	Mining Data Streams: sampling, filtering (e.g. bloom filters), sketch algorithms	<a href="http://infolab.stanford.edu/~ullman/mmds/ch4.pdf">http://infolab.stanford.edu/~ullman/mmds/ch4.pdf</a> (MMDS book Ch. 4)



3	Sketch Algorithms Cont. (Count-Min and Heavy Hitters)	Notes from <a href="http://theory.stanford.edu/~tim/s15/l12.pdf">http://theory.stanford.edu/~tim/s15/l12.pdf</a>
4	Mining Social Graphs: Intro to Social Networks as Graphs, Clustering (distance measures, Betweenness and Girvan-Newman alg.)	<a href="http://infolab.stanford.edu/~ullman/mmds/ch10n.pdf">http://infolab.stanford.edu/~ullman/mmds/ch10n.pdf</a> (MMDS book Ch. 10)
5	Mining Social Graphs Cont.: Community detection and Graph Partitioning (Finding Clique, Bipartite Graphs, Partitioning)	NA
6	Data Science lifecycle & Exploratory Data Analysis & Ethics of Big Data	<a href="http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf">http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf</a> (EDA book Ch. 4-7)  Voosen, P., Big-Data Scientists Face Ethical Challenges After Facebook Study. The Chronicle of Higher Education. Retrieved from <a href="https://www.chronicle.com/article/Big-Data-Scientists-Face/150871">https://www.chronicle.com/article/Big-Data-Scientists-Face/150871</a>
7	Data Visualization (including topics such as dimensionality reduction, tSNE)	(t-SNE) <a href="https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf">https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf</a>  (UMAP) <a href="https://arxiv.org/pdf/1802.03426.pdf">https://arxiv.org/pdf/1802.03426.pdf</a>
8	Data cleaning techniques	<a href="http://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf">http://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf</a>
9	Data Integration methods & ETL (schema matching, record-linkage, data fusion)	<a href="#">Big Data Integration</a> by Xin Luna Dong and Divesh Srivastava
10	Data Integration methods cont.	NA

## **CS 252B/EE 251B: Fundamentals of Data Science**

### **Spring 2021**

**Instructor:** Samet Oymak, Christian Shelton

**Contact Info:** [oymak@ece.ucr.edu](mailto:oymak@ece.ucr.edu), [cshelton@cs.ucr.edu](mailto:cshelton@cs.ucr.edu)

#### **Credits and type**

4.0 Units

Lecture: 3 hours

Research (outside): 3 hours

#### **Course Information**

##### **A. Course Description**

Explores theoretical tools in data science and their applications in data data science. Introduces and motivates statistical and computational viewpoints on data analysis. Topics include the manipulation of data as vectors, drawing inferences from data as distributions, and quantifying data uncertainty for data analysis. Also includes in-class and homework exercises on practical applications of these theoretical data science tools.

**B. Prerequisite(s)** Math 010A, Math 031 or EE020, CS100, Stat 155 or EE114 or equivalent, or permission by instructor

#### **Syllabus**

##### **Week 1**

Data as a vector I: motivation for linear algebra in data science, norms of vectors and matrices, eigenvalues and eigenvectors, fundamental subspaces

##### **Week 2**

Data as a vector II: Hermitian and positive semidefinite matrices, singular values, QR decomposition, principal component analysis (PCA), low-rank approximation

##### **Week 3**

Data analysis with linear algebra: least-squares, pseudo-inverse, condition number, ridge regression, in-class exercise on MNIST dataset and PCA

##### **Week 4**

Data as a distribution I: motivation for statistics and probability in data science, the randomness in data, random variable, conditional probability, expectation, variance, moments

##### **Week 5**

## Proposal for M.S. Degree in Computational Data Science

Data as a distribution II: covariance matrices, correlation coefficient, data normalization, multivariate Gaussians, law of large numbers, in-class exercise on analyzing covariance matrices on the Adult dataset

### **Week 6**

Inference with data: Parameter estimation, unbiased estimator, bias-variance decomposition, maximum likelihood estimator (MLE), maximum a posteriori estimation (MAP), log likelihood

### **Week 7**

Applications of Estimation: in-class exercises on MLE in clinical data, minimum mean-square error (MMSE), prediction with least-squares, coefficient of determination, in-class exercise on MMSE in time series prediction

### **Week 8**

Quantifying uncertainty with data: hypothesis testing, confidence intervals, p-value, Student's t-test, bootstrapping, in-class exercise on hypothesis testing on the movie ratings

### **Week 9**

Optimization with data: the role of data in modern optimization problems, loss functions, convexity, gradient, in-class exercise on gradient descent and least-squares on the Adult dataset

### **Week 10**

Overflow: Finish the material from earlier weeks or practice for the final exam.

### **Textbooks and Related Materials**

Recommended sources:

1. Textbook for linear algebra topics: Gilbert Strang, "Linear Algebra and Learning from Data", Wellesley-Cambridge Press, 2019.
2. Textbook for statistical topics: Larry Wasserman, "All of Statistics: A Concise Course in Statistical Inference", Springer, 2013. [available as a free pdf from <https://link.springer.com/content/pdf/10.1007%2F978-0-387-21736-9.pdf>]

### **Grading** TBD

Participation 5%

HWs 40% (mix of coding projects and problem solving on paper)

Midterm 25%

Final 30%

## **CS/EE 279 : Capstone Project in Data Science**

### **Fall 2022**

**Instructor:** Mariam Salloum

**Contact Information:** [msalloum@cs.ucr.edu](mailto:msalloum@cs.ucr.edu)

#### **Credits/Type**

4.0 Units

Lecture: 3 hours

Research (outside): 3 hour

#### **Short Description (<= 50 words)**

Covers combining technical, analytic, and interpretive skills to design and execute a large-scale data science capstone project that has a focus on real-world applications. Provides an opportunity to integrate all of the core skills and concepts learned throughout the program and prepares students for long-term professional success in the field. Emphasizes collaboration and communication in both written and oral form.

**Prerequisites:** Enrollment in Master in Computational Data Science.

**Co-requisites:** CS 252A/EE 251A, CS 252B/EE 251B, CS224, CS226, CS235, CS236.

#### **Course Objectives**

At the end of this course, students will be able to demonstrate their knowledge, skills and abilities to develop and execute a data science project using real-world data and effectively communicate their results to a technical and non-technical audience.

Students will be able to:

- Formulate a research question, problem or hypothesis that can be answered or tested using real-world data;
- Collect and manage data to devise solutions to their research question, problem or hypothesis;
- Select, apply and evaluate models, tools and methods to address their research question, problem or hypothesis. This includes building an end-to-end analysis pipeline covering data sourcing, cleaning/preparation, integration and transformation, and visualization;
- Interpret and assess their results and evaluate the limitations of their findings;
- Prepare a professional report of their work and effectively communicate their findings to a technical and non-technical audience.

## **Grading**

Students will work on a quarter-long project in teams of 2-3 students. The grading rubric is focused on group and individual project representations, project report, and a final web-based deliverable. In addition to these assignments, students are evaluated based on their participation in class discussions, and by their group-mates based on contributions to the group.

- 5% - Class participation (class discussions) and weekly meetings with course instructor
- 50% - Project Deliverables
  - Proposal (due Week 2) - Project proposal
  - Phase 1 (due Week 4) - Code and status report
  - Phase 2 (due Week 7) - Code and status report
  - Phase 3 (due Finals Week) - Code, and final report
- 35% - In-class presentations - instructor evaluation and peer feedback on presentations
- 10% - Web-based final deliverable

## **Readings**

There is no textbook for this course. Readings are drawn from various relevant books, articles and academic papers that are available online.

## **Schedule**

### **Week 1 - Introduction**

#### ***Topics***

- Reviewing the data science life-cycle
- Case studies of organizations using “big data” effectively
- Project and group selection

#### ***Required Readings***

- Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins and Nina Kruschwitz, Big data, analytics, and the path from insights to value. MIT Sloan Management Review (2010).  
<https://sloanreview.mit.edu/projects/analytics-the-new-path-to-value/>
- Robert, Christian. 2020. “The 9 Pitfalls of Data Science.” *Chance* 33. [UCR Library]
- (optional) Press. G. (2013). A very short history of data science. Forbes. Retrieved from  
<https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#7cfe337655cf>

## Week 2 - Data Science Applications

### *Topics*

- Data anonymity
- Selected readings from DS applications, focused on areas such as Social Media Analysis, Social and Information Networks, Healthcare and Medicine

### *Required Readings*

- Anna Monreale, Gennady Andrienko, Natalia Andrienko, Fosca Giannotti, Dino Pedreschi, Salvatore Rinzivillo, Stefan Wrobel. Movement Data Anonymity through Generalization, Transactions on Data Privacy, 2010. <http://www.tdp.cat/issues/tdp.a045a10.pdf>
- Selected readings from papers highlighted in <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9912/9874>
- (optional) Voosen, P., Big-Data Scientists Face Ethical Challenges After Facebook Study. The Chronicle of Higher Education. Retrieved from <https://www.chronicle.com/article/Big-Data-Scientists-Face/150871>

## Week 3 - Data Science Ethics

### *Topics*

- Skills for collecting, storing, sharing and analyzing data derived from human subjects including data used in algorithms and examining ethical implications.

### *Required Readings*

- 1 - Data Skeptic
  - O'Neil, Cathy. (2013) On being a data skeptic. p. 1-19. Sebastopol, CA: O'Reilly Media. <http://www.oreilly.com/data/free/files/being-a-data-skeptic.pdf>
  - boyd, danah. (2017)
  - Ethical side effects of the publish or perish system: p-hacking and small sample size Gelman, Andrew and Loken, Eric. (2014) The statistical crisis in science. American Scientist. Accessed online: <https://www.americanscientist.org/issues/id.16259,y.2014,no.6,content.true,page.1,css.pri nt/issue.aspx>
- 2- Data Sharing / Ethics
  - Libby Bishop (2017). Big data and data sharing: Ethical issues. UK Data Service, UK Data Archive. Accessed online: [https://bigdata.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing\\_ethicalissues.pdf](https://bigdata.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethicalissues.pdf)
- 3- Building fair systems/ Ethics
  - Toward accountability: Data, Fairness, Algorithms, Consequences. Data and Society: Points. [blog post] Accessed online: <https://points.datasociety.net/toward-accountability-6096e38878f0>

- Crawford, Kate. (2013, April 1) The hidden biases in big data. Harvard Business Review. <https://hbr.org/2013/04/the-hidden-biases-in-big-data> Lerman, Jonas. (2013, September)
- Big data and its exclusions. Stanford Law Review. Accessed online: <https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusion>

#### **Week 4 - Group Presentations I**

##### ***Topics***

- Group Presentations

##### ***Required Readings***

- NA

#### **Week 5 - Communication and Storytelling**

##### ***Topics***

- Power of storytelling and narrative
- Tactics for presenting and sharing information

##### ***Required Readings***

- Selected readings from [Interactive storytelling : 7th International Conference on Interactive Digital Storytelling, ICIDS 2014, Singapore, Singapore, November 3-6, 2014 : proceedings](#)

#### **Week 6 - Data Visualizations**

##### ***Topics***

- How people and organizations process information and make decisions
- Use of data visualization for communication

##### ***Required Readings***

- Selected readings from [2019 IEEE Visualization in Data Science \(VDS\)](#) IEEE Visualization in Data Science (Conference) (2019 : Vancouver, B.C.)
- Gavett, G. (2014) How data visualization answered of retail's most vexing questions. Harvard Business Review. Retrieved from <https://hbr.org/2014/05/how-data-visualization-answered-one-of-retails-most-vexing-questions>

#### **Week 7 - Group Presentations II**

##### ***Topics***

- Group Presentations

***Required Readings***

- NA

**Week 8 - New trends / topics in Data Science**

***Topics***

- Highlight current research work in data science

***Required Readings***

- Selected readings from KDD, ICML, VLDB, IEEE Big Data

**Week 9 - Guest presentations**

***Topics***

- Guest speaker will discuss their experience in industry

***Required Readings***

- NA

**Week 10 - Final Group Presentations and Deliverables**

***Topics***

- Prepare for final in-class group presentations
- Deliver final presentations and submit project deliverables

***Required Readings***

- None



## APPENDIX C: LETTERS OF SUPPORT

Letters from BCOE Dean, CSE and ECE Department Chairs

Letters from other UCs and Universities:

- *Ian Davidson*; Professor of Computer Science; Chancellor's Fellow; member, Provost and Chancellor's committees on Data Science; **University of California, Davis**
- *Erik B. Sudderth*; Professor of Computer Science and Statistics; Director, Center for Machine Learning and Intelligent Systems; **University of California, Irvine**
- *Wei Wang*; Leonard Kleinrock Professor in Computer Science and Computational Medicine; Director, Scalable Analytics Institute; **University of California, Los Angeles**
- *Ambuj K. Singh*; Professor of Computer Science; Director of Data Science Initiative; **University of California, Santa Barbara**.
- *Lise Getoor*; Baskin Chair Professor, Computer Science & Engineering; Director, Data, Discovery and Decision (D3) Data Science Research Center; **University of California, Santa Cruz**
- *Rajesh Gupta*; Professor of Computer Science and Engineering; Director, Halicioğlu Data Science Institute; **University of California, San Diego**
- *Michael C. Yip*; Associate Professor of Electrical and Computer Engineering; Curriculum Advisor on the M.S. degree on Machine Learning and Data Science; **University of California, San Diego**
- *Juliana Freire*; Professor, Computer Science and Engineering and Data Science; founding member and former Graduate Director, NYU Center for Data Science; **New York University**
- *Christos Faloutsos*; Fredkin Professor of Computer Science; faculty member, MS of Computational Data Science; **Carnegie Mellon University**

Letters from Local Government, Organizations and Industry:

- *Jennifer Claar*, PhD; Managing Director, **Department of Public Social Services, County of Riverside**
- *Brandon Davis*, PhD; President, **Council for the Advancement of Black Engineers**
- *Doran J. Barnes*; Chief Executive Officer, **Foothill Transit**
- *Zaid J. Towfic*, PhD; Flight Communications Systems, **NASA Jet Propulsion Laboratory**
- *Geri Miller*; Education Sector Director, Global Business Development, **ESRI Inc.**

November 12, 2021

To whom it may concern:

I am writing this letter in enthusiastic support for the enclosed proposal to establish a Master of Science degree program in Computational Data Science. This program will be jointly administered within BCOE by the departments of Electrical and Computer Engineering and Computer Science and Engineering. I have had detailed conversations with Professor Tsotras and the program committee and fully support the academic program and administrative structure. I commit to working with them to insure the program's success.

This program will help address the critical and documented shortage of college graduates educated in Computational Data Science and the critical interpretation and analysis of large datasets. We expect students attracted to this program to come from a variety of backgrounds and other interests, increasing the diversity among Engineering students, and those in computational fields in particular.

As the program grows, it will require support staff and office space. BCOE will provide the resources for both.

The Bourns College of Engineering looks forward to launching this MS Computational Data Science degree program. It is an important part of keeping our curriculum current and educating our students.

Sincerely,



Prof. Christopher S. Lynch  
William R. Johnson Jr. Family Chair  
Dean, Bourns College of Engineering  
University of California, Riverside



## Computer Science and Engineering

Walid A. Najjar  
Professor and Chair  
Department of Computer Science and Engineering  
351 Winston Chung Hall Riverside, CA 92521  
Tel 951.827.5639 Fax 951.827.4643  
[najjar@cs.ucr.edu](mailto:najjar@cs.ucr.edu) [www.cs.ucr.edu/~najjar](http://www.cs.ucr.edu/~najjar)

November 11, 2021

To Whom It May Concern:

This letter is in strong support for the proposed Master of Science (MS) program in Computational Data Science at UCR, to be jointly offered by the Departments of Computer Science and Engineering and Electrical and Computer Engineering.

Data has become ubiquitous in everyday life revolutionizing science and industries alike. Data Science has grown out of the need to study the data itself and in particular, how to manage, process and interpret data. Tools originating from data science are now becoming indispensable in today's science, technology, and business, fueling the demand for data scientists. Recognizing this need, our department has taken the initiative to develop research and educational programs related to Data Science at UCR.

In collaboration with other departments on campus, an online MS program in Data Science is already being offered. This fall we welcomed the second incoming class of the new undergraduate program in Data Science (offered in collaboration with the Statistics Department). Recently the Data Science Center has been established that includes multiple newly hired faculty members, and has been given designated space in the new MRB building. Given the very strong faculty expertise in computational methods of Data Science (AI, Data Mining, Big Data, Machine Learning, Visualization etc.) between the Departments of Computer Science & Engineering and Electrical and Computer Engineering, creating a state-supported MS program focusing in Computational Data Science is the next step in this endeavor.

This program will address critical and documented shortage of highly trained college graduates with an advanced degree in Data Science, in industry, government, and academia. The CSE Department enthusiastically supports the creation of the Computational Data Science MS program and is fully committed to providing necessary resources within its capabilities for the instruction and advising of its students.

A handwritten signature in blue ink that reads "Walid A. Najjar".

Walid A. Najjar

Professor and Chair  
Department of Computer Science and Engineering  
Bourns College of Engineering  
University of California Riverside

November 12, 2021

Dear Members of the Academic Senate:

It is my pleasure to provide the strongest possible support for the MS in Computational Data Science program. This program will be housed in the Bourns College of Engineering, and is cross-disciplinary, across the departments of Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE). It will draw upon courses from the existing programs from the departments, including three new cross-listed courses.

Computational Data Science is strategically and technically a very important area that studies how to obtain insight and information from the analysis of large collections of data. As data has become ubiquitous in everyday life, it impacts every profession, including manufacturing, logistics, health care, public safety, and the military. Data is also important in all aspects of science and engineering. The proposed MS in Computational Data Science is a comprehensive program studying how data can be collected, transformed, analyzed, and used to solve problems across many application areas. Students will acquire the cross-disciplinary breadth required for this important and emerging field and can focus, through electives, on specific areas of interest. The proposed program does so at very little expense, since the teaching and research infrastructure are already in place.

ECE expects to interact extensively with the proposed MS in Computational Data Science program by participating in teaching the required and elective courses, in data science research and the mentoring of students through projects and advising, and in helping with the program administration. The program will contribute in a great many positive ways to the ECE department.

In summary, I am extremely supportive of this program and believe it will greatly benefit the students and will help raise UCR's profile. Please do not hesitate to contact me should there be any questions. Sincerely,



Ertem Tuncel  
Professor and Chair,  
[Electrical and Computer Engineering](#)  
University of California, Riverside



DEPARTMENT OF COMPUTER SCIENCE  
ONE SHIELDS AVENUE  
DAVIS, CALIFORNIA 95616-5270, USA

TELEPHONE +1 530 752 7004  
FAX +1 530 752 4767

November 12, 2021

Ian Davidson  
Department of Computer Science  
University of California - Davis  
Davis, CA 95616

To Whom It May Concern,

I am a full Professor of Computer Science at the University of California – Davis. I served on both our Provost and Chancellor’s committees on Data Science. For the past two decades I have published extensively in the fields of Artificial Intelligence, Data Mining and Machine Learning.

I am writing this letter in enthusiastic support of the proposed MS degree in ”Computational Data Science” as put forward by the departments of Computer Science and Engineering and Electrical and Computer Engineering at UCR.

In my review of the proposal, I was very encouraged to see that the MS program is offering a comprehensive, rigorous, and well-thought-out curriculum that broadly covers computational aspects of data science. Data science is a multidisciplinary field, and its computational aspects span a number of areas, such as machine learning, data mining, and artificial intelligence, which are in very strong demand by both industry and government. The proposed MS program will equip students with the necessary skills to be competitive in the job market and thrive in relevant positions.

I know professionally several of the faculty at Riverside including Professor Keogh and Papalexakis. Not only are they outstanding world leaders in their research fields, but I’ve heard them given talks and tutorials which demonstrated they are outstanding communicators. The proposed program perfectly leverages the collective expertise of computationally intensive data science faculty in both departments, who are at the forefront of the field and will undoubtedly keep the program up to date and relevant to the job market and the research community.

The program’s curriculum is comparable with computational data science MS programs already offered by top institutions in the nation, such as Carnegie Mellon University and UC San Diego. I was glad to see that the curriculum includes several new courses that are unique to data science. ECS252A and B in particular seem like well thought out foundational courses that will help meld the other offerings together. A capstone course is essential to ground the material and I was happy to see its inclusion.

I believe that the MS in Computational Data Science will further solidify UCR’s rising reputation in data science. Moreover, as computational data science jobs are among the highest-paying in today’s market, it will amplify UCR’s outstanding record in social mobility. I regularly am in Silicon Valley giving talks to companies and a common request I get is to send well rounded students. I think the students this program will be ideal for these companies and I am delighted

that UCR is fulfilling this vital need.

Sincerely,

Sincerely.

A handwritten signature in blue ink that reads "Ian Davidson". The signature is written in a cursive, flowing style. There is a small horizontal line at the end of the signature.

Professor Ian Davidson  
Computer Science  
University of California Davis  
(530) 601 0385  
davidson@cs.ucdavis.edu

# UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

SCHOOL OF INFORMATION AND COMPUTER SCIENCES  
DEPARTMENT OF COMPUTER SCIENCE

IRVINE, CALIFORNIA 92697-3425  
Phone (949) 824-4478 Fax (949) 824-4056

November 4, 2021

Professor Vassilis Tsotras  
Department of Computer Science & Engineering  
University of California  
Riverside, CA 92521

Professor Tsotras,

Your initiative to create a new Masters program in Computational Data Science is thoughtfully designed and thoroughly supported. It will be in high demand and provide valuable education and training for the needed growth in data scientists with joint strength in computational methods and analysis techniques. The various approaches taken by the proposal to improve diversity are very commendable, including leveraging UCR's recent grant from the NSF Harnessing the Data Revolution Data Science Corps.

At UC Irvine I am a Chancellor's Fellow, Director of the UCI Center for Machine Learning and Intelligent Systems, and Director of the HPI Research Center in Machine Learning and Data Science at UCI. I joined the faculty at UCI in 2017, after eight years on the faculty in the Brown University Department of Computer Science. I have seen the student interest and employer demand for data scientists grow dramatically over the past decade, and this proposal is well positioned to leverage these trends.

The curriculum focuses on the computational aspects of data science: how to store and process the volumes of data necessary at the speeds necessary, as well as the principles (drawn from machine learning, artificial intelligence, databases, and information retrieval) underlying effective and robust data analyses. The computationally focused curriculum is well designed to provide a solid education in only a little more than a year, and should be very attractive to those with an undergraduate degree in computer science or electrical engineering, who want to better understand how their training in computer software and hardware may be applied to state-of-the-art data analysis. Note that the topics covered by a Computational Data Science degree are very different from the existing UCR MS in Business Analytics, and I expect there to be a large and distinct set of prospective students eager to learn about them. The program is also distinct and complementary from UC Irvine's own Professional Master of Data Science, which is a self-supporting graduate professional degree program (SSGPDP) with greater emphasis on connections to statistics and machine learning.

You have an excellent set of faculty to support this program, spanning spatial databases to data mining to machine learning. I expect that experience in teaching and administering MS programs in Computer Science and Electrical Engineering will allow for a smooth start to this new program.

With a well-designed curriculum addressing an important topic, that will receive demand from diverse sources, this proposal for a MS in Computational Data Science has my full support.

Yours sincerely,

A handwritten signature in black ink that reads "Erik B. Sudderth". The script is fluid and cursive, with the first letters of the first and last names being capitalized and prominent.

Erik B. Sudderth  
Professor of Computer Science and Statistics  
University of California, Irvine

*Erik B. Sudderth is Professor of Computer Science and Statistics, and Chancellor's Fellow, at the University of California, Irvine. He directs the UC Irvine Center for Machine Learning and Intelligent Systems, as well as the HPI Research Center in Machine Learning and Data Science at UC Irvine. His research interests include probabilistic graphical models and probabilistic programming, nonparametric Bayesian methods for weakly supervised learning, and applications of statistical machine learning in computer vision and the sciences. Erik was previously an Associate Professor of Computer Science at Brown University, and a postdoctoral scholar at the University of California, Berkeley. He received the Bachelor's degree (summa cum laude, 1999) in Electrical Engineering from the University of California, San Diego, and the Master's and Ph.D. degrees (2006) in Electrical Engineering & Computer Science from the Massachusetts Institute of Technology. He is an associate editor for IEEE Transactions on Pattern Analysis & Machine Intelligence, an action editor for the Journal of Machine Learning Research, and a regular area chair for the top conferences in machine learning and computer vision. Erik received an NSF CAREER award, the ISBA Mitchell Prize, and was named one of "AI's 10 to Watch" by IEEE Intelligent Systems Magazine.*



---

Computer Science Department

November 10, 2021

To Whom it May Concern:

I am writing this letter to enthusiastically endorse the proposal for the MS in Computational Data Science prepared jointly by the Department of Computer Science and Engineering and the Department of Electrical and Computer Engineering at UCR.

I am the Leonard Kleinrock Chair Professor in Computer Science and Computational Medicine at University of California, Los Angeles and the director of the Scalable Analytics Institute (ScAi). I have been 20+ years of experience in computational data science.

First, I need to emphasize the relevance and timing of such a program. Companies of all types and industries depend more on data science for their decisions and need to hire people who are able to work and analyze big data. The focus on the Computational part of Data Science, differentiates this program for other relevant MS programs in Data Science and/or Business Analytics. The new program will provide its students with an in-depth understanding of the data lifecycle, including data collection, data cleaning, data integration, data management, and data visualization, as well as the theories and techniques necessary for data analysis from data mining, machine learning, information retrieval, and artificial intelligence. The demand for the graduates of the MS in Computational Data Science will remain strong for the foreseeable future, as more industries, businesses and government entities continue to move to the direction of big data management and analytics.

Graduates from this 5-quarter MS program will be well-prepared either to a career as a Data Scientist or for a transition to a PhD. It is also important to mention the many approaches that the two departments have and will undertake to attract women and underrepresented minorities to the new MS in Computational Data Science. This is of special significance to UCR given its current status as the top social mobility university. Another advantage of the new program is its synergy with the summer Bridge programs which would allow students from non-computing backgrounds to build the needed preparation before entering the MS in Computational Data Science and still finish the MS within 2 years.

UC Riverside has an exceptional cadre of faculty across Computer Science and Electrical Engineering, with strong expertise in Artificial Intelligence, Machine Learning, Data Mining, Time Series, Big Data and Visualization, etc. Such expertise makes the program faculty uniquely qualified in running the MS in Computational Data Science.

The curriculum is carefully designed to build a thorough background and the skills needed for a professional career in Computational Data Science. The details of the program's administration, admission, evaluation, and resources, are well thought out and, in my opinion, this program will be very successful and competitive in the growing job market in Data Science.

To conclude, this is a strong proposal for a timely degree in MS in Computational Data Science, with support from a well-qualified group of faculty members. I have no doubt that the program will attract strong interest and that it will keep growing over the years. I thus enthusiastically endorse this proposal.

Sincerely,



Wei Wang, Ph.D.

Leonard Kleinrock Professor in Computer Science

Director, Scalable Analytics Institute (SCAI)

University of California, Los Angeles

[www.cs.ucla.edu/~weiwang](http://www.cs.ucla.edu/~weiwang)



DEPARTMENT OF COMPUTER SCIENCE  
SANTA BARBARA, CALIFORNIA 93106-5110

PHONE: (805) 893-3236  
EMAIL: [ambuj@cs.ucsb.edu](mailto:ambuj@cs.ucsb.edu)  
December 15, 2020

Professor Vassilis Tsotras  
Department of Computer Science and Engineering  
Bourns College of Engineering  
University of California  
Riverside, CA 92521

Dear Vassilis:

I am writing to strongly support the proposal by the Department of Computer Science and Engineering and the Department of Electrical and Computer Engineering to create a MS in Computational Data Science at the University of California, Riverside. I believe that this proposal addresses current needs in Industry and Academia for students with specialized training in the broad disciplines of artificial intelligence, machine learning, data mining, big data management among others, which blend together in the Data Science discipline. The interdepartmental program proposed at the University of California, Riverside, will provide the students with the required technical knowledge to pursue their careers in the field of Data Science. The proposed MS degree with its focus on the Computational part of Data Science will distinguish this program among other Data Science related degrees, and will likely inspire other institutions to launch similar programs in the near future.

I have been a faculty at the University of California, Santa Barbara, where I lead the Data Science Initiative. This initiative is implementing a distributed curriculum across departments. The current focus is on the undergraduate end but we do want to consider the development of graduate curriculum too. With that in mind, I like the structure of your program and look forward to seeing the progress. My personal research is on machine learning on networks, data management, computational social science, and network science. As such I recognize and understand the focus areas of your proposed program.

The need for students trained in the area of Data Science is huge and growing. This is a great opportunity for students from underrepresented groups to obtain a MS degree in a very popular field that will enhance both the financial aspects and success of their future careers. The proposed program will make special efforts to attract women and underrepresented minorities in Data Science and computing in general. As a university that values the diversity of its student population, UCR will benefit greatly from offering the MS in Computational Data Science to its students.

The proposed curriculum is well-thought-out and balanced. It focuses on students with experience in a quantitative field and basic undergraduate knowledge in programming, software engineering, algorithms, and introductory statistics. Students will need to complete a number of required classes from the two participating Departments, a capstone project, as well as five elective courses. The extensive set of elective courses would allow students to further focus on specific subjects. The program's structure guarantees that graduating students will have the required background and skills, but also the knowledge and flexibility to work in Data Science related positions. The program's duration is a little over a year (4 quarters) which will

be attractive to future students.

The participating Departments at the University of California, Riverside, are well positioned to initiate, administer, and grow the proposed MS program. The majority of the courses needed for the new program are currently being offered, thus minimizing the time, effort and risks to launch the new program. The organizational structure of the program is also very reasonable and effective, with Director and Co-Directors roles that will oversee course offerings, scheduling, and other practical matters.

In summary, I strongly support this proposal and I believe that this new program will attract numerous high quality students soon after its launch. Please do not hesitate to contact me should you need further information.

Sincerely,

A handwritten signature in black ink that reads "Ambuj Kumar Singh". The signature is written in a cursive, flowing style.

Ambuj K. Singh  
Professor  
Dept. of Computer Science



LISE GETOOR  
JACK BASKIN SCHOOL OF ENGINEERING  
COMPUTER SCIENCE & ENGINEERING DEPARTMENT  
ENGINEERING 2 BUILDING  
SANTA CRUZ, CALIFORNIA 95064

(831)459-1489  
(831)459-4829 FAX  
GETOOR@UCSC.EDU

November 10, 2021

Professor Vassilis Tsotras  
Department of Computer Science & Engineering  
University of California  
Riverside, CA 92521

Dear Professor Tsotras,

I am writing to offer my support for the proposed new Masters program in Computational Data Science at UC Riverside. As the Director of the UC Santa Cruz D3 Data Science Research Center and PI on NSF TRIPODS Institute for Foundations of Data Science, I reviewed the proposal with interest.

The program shared with me is well targeted for students with a computational background. In just a few quarters, the curriculum builds on students' undergraduate computational expertise to provide them with understanding of the tools and techniques of data science. The program nicely spans the range of computational efforts in data science from databases to data mining. Your faculty, across ECE and CSE, are well qualified to deliver these courses and provide a quality degree.

Computational data scientists are in high demand. The proposed major takes on a variety of special efforts to broaden participation and to attract minorities and women in this important field. The students' education in this fast-growing field will allow them to elevate their careers.

In summary, this is a strong, timely, well-targeted proposal supported by well-qualified faculty. It has my full support.

Sincerely,

A handwritten signature in black ink, appearing to read "Lise Getoor".

Lise Getoor  
Director, Data, Discovery and Decision (D3) Data Science Research Center  
Baskin Chair Professor, Computer Science & Engineering Department and Engineering, UC Santa Cruz



*Bio: Lise Getoor is the Baskin Chair Professor in the Computer Science & Engineering Department at UC Santa Cruz, and founding Director of the UC Santa Cruz D3 Data Science Research Center. Her research areas include machine learning and reasoning under uncertainty; in addition, she works in data management, visual analytics and computational social science. She has over 250 publications and extensive experience with machine learning and probabilistic modeling methods for graph and network data. She is a Fellow of the Association for Computing Machinery (ACM), the Association for Artificial Intelligence (AAAI), and the Institute of Electrical and Electronic Engineers (IEEE). She has served as an elected board member of the International Machine Learning Society, has served on the board of the Computing Research Association (CRA), has served as Machine Learning Journal Action Editor, Associate Editor for the ACM Transactions of Knowledge Discovery from Data, JAIR Associate Editor, and on the AAAI Executive Council. She was co-chair for ICML, and has served on the PC of many conferences including the senior PC of AAAI, ICML, KDD, UAI, WSDM and the PC of SIGMOD, VLDB, and WWW. She is a recipient of an NSF Career Award and thirteen best paper and best student paper awards. She was selected to give the UC Santa Cruz Faculty Research Lecture 2018-19, one of the highest recognitions given to UC faculty. In 2019, she was selected as a Distinguished Alumna of the UC Santa Barbara Computer Science Department and she also received the UCSC Women in Science & Engineering (WISE) award, for her efforts mentoring women in computer science. She has recently given keynotes and talks on “Responsible Data Science” at SIGMOD 19, IEEE Big Data 19, ScaledML 19, and talks on Ethics and Data science at NSF workshops on Teaching Data Science Ethics, and the Global Forum on AI for Humanity. She received her PhD from Stanford University in 2001, her MS from UC Berkeley, and her BS from UC Santa Barbara, and was a professor at the University of Maryland, College Park from 2001-2013.*



November 29<sup>th</sup>, 2021

Professor Vassilis Tsotras  
Department of Computer Science & Engineering  
University of California - Riverside  
Riverside, CA 92521

Dear Vassilis,

Thank you for sharing the draft of the proposed MS in Computational Data Science at UC Riverside. The new program will train students in the computational part of Data Science offering a variety of state-of-the-art courses in Machine Learning, Big Data Management, Data Mining, Deep Learning and Artificial Intelligence. In that context, I find that the curriculum is well-thought. The collaborating Department of Computer Science and Engineering and the Department of Electrical and Computer Engineering certainly have faculty with expertise in these areas, thus making it possible to launch a viable program with accessible courses and instructors to the students.

My assessment on this program is informed by my appointment and experience in leading data science at UC San Diego, first as chair of the Computer Science and Engineering (CSE) department and currently as the founding director of the Halicioğlu Data Science Institute (HDSI) at UC San Diego. As chair of CSE, I had led the design and launch of our popular undergraduate major and minor in Data Science in addition to our professional masters program in Data Science and Engineering. As a researcher, I work on cyber-physical systems with a focus on sensor data organization and its use in optimization and analytics. I have led several large scale projects including the National Science Foundation (NSF) Expeditions on Variability, and Defense Advanced Research Projects Agency (DARPA) projects under Data Intensive Systems (DIS) and Circuit Realization at Faster Timescales (CRAFT) programs. As the director of the Halicioğlu Data Science Institute I have strong interest in, and experience with, Data Science related educational programs and efforts.

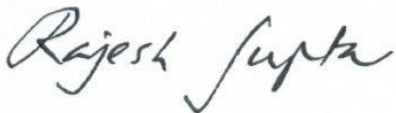
Overall, I find that the proposed program is timely and I expect that your graduates will be sought after by the related industry and government. Data Science related jobs are on the rise and offer a great and secure career path for our students. It is my understanding that this new M.S. will be a state-supported program with in-class lecturing. Your program is tailored towards attracting recent B.S. graduates majoring in a quantitative field, with basic knowledge in programming, data structures/algorithms and statistics. As such this program will be different from other professional M.S. degrees in Data Science or

Business Analytics. Further, its focus on Computational Data Science, makes it a unique offering for UCR among other UCs.

Finally, I am encouraged by the proposal's focus and methods on improving diversity. Data Science as a field offers a large potential in attracting minorities and underrepresented students in the computing field. The proposed program will offer your students secure career paths in Data Science and is thus complementing UCR's status as a leader in social mobility.

In summary, I find your proposal timely and well designed. The focus on Computational Data Science makes your program unique. I expect this new MS to be popular amongst students. I would like to express my enthusiastic support for the proposed M.S. in Computational Data Science at UC Riverside.

Sincerely,

A handwritten signature in black ink that reads "Rajesh Gupta". The signature is written in a cursive, flowing style.

Rajesh Gupta, PhD  
Director, Halıcıoğlu Data Science Institute





MICHAEL C. YIP  
ASSOCIATE PROFESSOR  
DEPT. ELECTRICAL & COMPUTER ENGINEERING  
9500 GILMAN DRIVE, MC 0407, LA JOLLA, CA 92093

PHONE: 858-822-4778  
E-MAIL: [yip@ucsd.edu](mailto:yip@ucsd.edu)

December 3, 2021

Professor Vassilis Tsotras  
Department of Computer Science & Engineering  
University of California - Riverside  
Riverside, CA 92521

Dear Professor Tsotras,

I have examined the proposal that you recently shared with me about the creation of a new M.S. degree in Computational Data Science at UC Riverside. I am writing to express my strong support for this new M.S. program. Given the large need by industry and government for workforce with Data Science expertise, the new M.S. is timely and I expect that it will be in high demand among students. The proposed program's focus on Computational Data Science differentiates it from other existing offerings in Data Science education.

I am an Associate Professor of Electrical and Computer Engineering at UC San Diego and Director of the Advanced Robotics and Controls Laboratory (ARCLab). I am also directing the Medical Robotics Collab in the Contextual Robotics Institute at UCSD. My research focuses on solving problems in data-efficient and computationally efficient robot control and motion planning through the use of various forms of learning representations, including deep learning and reinforcement learning strategies. My lab routinely publishes at the top conference for robotics, AI, and machine learning, and the work has been recognized through several best paper awards. I have been awarded the NSF CAREER award and the NIH Trailblazer Award, the Hellman Fellowship, a Distinguished Lecturer title with the IEEE Robotics and Automation Society.

Further, I serve as the Curriculum Advisor on the M.S. degree on Machine Learning and Data Science, offered by the Electrical and Computer Engineering Department at UCSD. Based on my research expertise and involvement with our own M.S. in Machine Learning and Data Science, I find the proposed program to be well designed, thoroughly covering the various aspects of Computational Data Science. This new interdepartmental degree will be offered by the Departments of Computer Science & Engineering and Electrical & Computer Engineering. The two departments are well-positioned to offer this program as their faculty include experts in all areas that comprise the computational part of Data Science (i.e., big data, visualization, spatial databases, scalable database management, data mining, artificial intelligence and machine learning).



It has been well established that Data Science has the potential to attract more females to computing. We have seen a modest increase in female enrollment (F-20.4%, M-72.1%, Undeclared 7.5%) over the other majors (F-15.5%, M-77.4%, Undeclared-7.1%). Notably, as an MLDS advisor, I have conversations with prospective female students in their consideration to enter or switch into the MLDS program, who perceive less barrier to entry into a Data Science career than their first Majors. Thus, the program's commitment and multiple approaches (including the bridge program) to increase diversity and attract students from underrepresented groups is commendable.

In summary, I enthusiastically support this timely program; it is well designed and will be highly sought after by students.

Sincerely,

A handwritten signature in cursive script that reads "Michael Yip".

Michael C. Yip, PhD

**NYU****TANDON SCHOOL  
OF ENGINEERING****Juliana Freire**  
*Professor of Computer Science and  
Data Science***Department of Computer Science/VIDA Center**  
370 Jay Street, 11<sup>th</sup> floor  
Brooklyn, NY 11201  
**P:** 646 997-4128  
juliana.freire@nyu.edu

November 16, 2021

*Re: Endorsement of the proposal for a MS in Computational Data Science*

Dear Members of the UCR Academic Senate Review Committee:

I am writing this letter I would like to express my enthusiastic endorsement of the new MS degree in Computational Data Science, proposed by the Departments of Electrical Engineering and Computer Science at UC Riverside. This is a timely degree and its focus on Computational Data Science fills an important gap in existing programs. I believe it will be very popular!

I am a Professor of Computer Science and Data Science at New York University. I was the NYU lead investigator for the Moore-Sloan Data Science Environment, a \$32.8 million grant awarded jointly to UW, NYU, and UC Berkeley. I am a founding member of the NYU Center for Data Science and served as its Graduate Director from 2014 through 2017. The overarching goal of my research is to develop methods and systems that enable a wide range of users to obtain trustworthy insights from data. This spans topics in large-scale data analysis and integration, visualization, machine learning, provenance management, and web information discovery, and different application areas, including urban analytics, predictive modeling, and computational reproducibility. I have co-authored over 200 technical papers (including 11 award-winning publications) and several open-source systems. According to Google Scholar, my h-index is 61 and my work has received over 15,900 citations. I am an ACM Fellow and a recipient of an NSF CAREER, two IBM Faculty awards, a Google Faculty Research award, and the 2020 ACM SIGMOD Contributions Award. I was elected chair of the ACM Special Interest Group on Management of Data (SIGMOD) and served as a council member of the Computing Research Association's Computing Community Consortium (CCC).

Data and computing have revolutionized science, industry, and government alike. In science, it has led to great advances and discoveries; companies are capitalizing on data, in fact, for many businesses, the main source of revenue comes from data; and increasingly, governments are collecting and using data to improve their operations, inform policies and decisions. The need to manage, analyze, and extract knowledge from data is pervasive. Professionals in different areas such as manufacturing, logistics, health care, public safety, and the military to name a few, from entry-level office workers to CFOs and CEOs, must be literate in data and computing to do their jobs effectively.

Data Science has emerged as a field that draws from diverse disciplines, including computer science, statistics, and mathematics, to develop methods and tools for managing, analyzing, modeling, and deriving insights from data. Fueled by the growing volumes of data, Data-Science-

related jobs have proliferated – there is a great demand for data scientists. According to Glassdoor, a recruiting site, Data Scientist is among the best jobs in the US for the last five years in a row, with a high median salary.

The proposed MS in Computational Data Science focuses on the foundational computing principles of data science, notably Machine Learning, Artificial Intelligence, Big Data, and Visualization. Overall, the program is well designed, consisting of six core courses that cover basic knowledge required in common data science tasks, a practical capstone project, and five electives, which give students flexibility to seek different specializations. At NYU, I have seen a growing number of students from STEM-related disciplines that want to learn and gain expertise in computer science (CS), but existing CS programs are not suitable for them. The proposed program provides a well thought out path, including bridge courses, that can accommodate such students.

The program faculty has expertise and strong research records in the areas covered by the program. The computational focus distinguishes the proposed program from other related programs. And the computational focus is on high demand among industry leaders including companies like Google, Microsoft, Oracle, Teradata, eBay, Amazon. I expect the new MS to be very popular among students and it will thus serve well the UCR community.

Data Science has an enormous potential for increasing diversity in computing. As a point of reference, at NYU, our MS in data science has attracted a much higher number of female student than our CS programs: in 2019, over 50% of the enrolled MS students were female. I see the goal of attracting under-represented minorities as a highlight of this program. The proposal outlines a detailed plan and strategies that follow a set of suggested ‘best-practices’ from a National Academies report on the recruitment, retention, and success of diverse student populations in STEM and computing disciplines. In addition, program faculty have a strong record in Broadening Participation in Computing efforts, and the various related grants they have received serve as evidence.

In summary, the proposed program is well designed, timely, and fills an important gap in Data Science education. I henceforth offer my very strong support to the new MS in Computational Data Science.

Sincerely,

A handwritten signature in blue ink that reads "Juliana Freire". The signature is fluid and cursive, with the first name and last name clearly distinguishable.

Juliana Freire, Ph.D.  
Professor, Computer Science and Engineering and Data Science  
New York University  
[juliana.freire@nyu.edu](mailto:juliana.freire@nyu.edu)



Christos Faloutsos  
Fredkin Professor of Computer Science  
Computer Science Department  
Cell#: 412-576.7932  
[christos-admin-support@cs.cmu.edu](mailto:christos-admin-support@cs.cmu.edu)

December 3, 2021

Professor Vassilis Tsotras  
Department of Computer Science & Engineering  
University of California  
Riverside, CA 92521

Dear Professor Tsotras:

I am writing this letter to express my strongest support for the proposed Master's degree in Computational Data Science at the University of California – Riverside.

As modern technology has enabled the creation and storage of increasing amounts of information, data volumes have exploded. It is estimated that Facebook users upload 10 million photos every hour; almost 90 percent of the data in the world was created in just the last two years. The wealth of data being collected and stored by these technologies can bring transformative benefits to businesses, local government and societies around the world—but only if we can interpret it. Data Science has emerged as a new scientific field to manage and analyze such Big Data.

Computational Data Science in particular focuses on the computational foundations of data science, providing an in-depth understanding of the algorithms and data structures for storing, manipulating, visualizing and learning from large data sets. It is thus well timed that the Departments of Computer Science and Engineering and the Department of Electrical and Computer Engineering at UCR have joined forces to offer the Master in Computational Data Science degree. The two departments have faculty that are experts in crucial areas for such a MS degree, including Machine Learning, Data Mining, Artificial Intelligence and Big Data Management. Graduates from the proposed program will be thus well equipped with the skills and knowledge needed to analyze and manage big data. I expect that your graduates will be in high demand and will become data scientists and project managers at companies like Amazon, Apple, Facebook, Google, LinkedIn, Twitter, Teradata and Yahoo.

I am in good position to comment on the proposed program because of: (i) my research interests in Data Science and, (ii) my experience from our MS of Computational Data Science at Carnegie Mellon University. I am the Fredkin Professor in Artificial Intelligence at the Computer Science Department in CMU and my research interests include large-scale data mining with emphasis on graphs and time sequences, anomaly detection, tensors, and fractals. I have received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions Award in ICDM 2006, the SIGKDD Innovations Award (2010), the PAKDD Distinguished Contributions Award (2018), 29 "best paper" awards (including 8 "test of time" awards), and four teaching awards. I am also one of the faculty affiliated with CMU's Master

of Computational Data Science. Similar to the proposed program at UCR, our MS focuses on teaching students with the skills and knowledge needed to develop the technology involved in the next generation of massive information system deployments and to analyze the data those systems generate. We have seen high interest among CMU students in the Computational Data Science MS. Our graduates typically find employment in top industry players in the field while some decide to pursue a Ph.D. degree.

Of particular importance are the efforts outlined in your proposal for increasing diversity and attracting underrepresented minorities to computing. Many of these efforts have emanated from grants that your faculty have received on broadening participation. Given UCR's prominence in social mobility, the proposed new MS will offer new career opportunities to the student population served at UCR.

In summary, the proposed MS in Computational Data Science is a well-designed program with a curriculum supported by an exceptionally strong faculty, offering first rate education in an emerging field. I am confident your program will attract many qualified students and I thus enthusiastically support it.

Yours sincerely

A handwritten signature in blue ink, appearing to read 'CFaloutsos', is written over a light blue rectangular background.

Christos Faloutsos

Fredkin Professor of Computer Science

Carnegie Mellon University



**Administrative Office**  
4060 County Circle Drive, Riverside, CA 92503  
951.358.3000 FAX: 951.358.3036  
[www.rivcodpss.org](http://www.rivcodpss.org)

**Sayori Baldwin, Assistant CEO - Human Services**  
**Jennifer Claar, PhD, Managing Director**

November 15, 2021

To Whom It May Concern:

I am writing to express my support for the proposal by the Department of Computer Science and Engineering and the Department of Electrical and Computer Engineering to create a new Master of Science program in Computational Data Science at the University of California, Riverside. The proposed program addresses a current and real need in the Inland Empire for workforce trained in collecting, cleaning, visualizing, and analyzing large data sets. The computational focus of this program is unique as it offers an in-depth training of computational skills required to manage and analyze big data, moreover, the diversity plan will likely attract more women and underrepresented minorities in Data Science and contribute to the diversity in the computing field as a whole.

At the County of Riverside, we have taken steps to leverage ML to draw insight from our data and better understand patterns in hiring, advancement, and retention. A unique and exciting feature of this program is the capstone project. We are currently working with Dr. Salloum and her Senior Design students at UCR on analyzing retention patterns for social workers at the County of Riverside, and this program offers further opportunities for industry to collaborate with students in the new MS. This is a great opportunity for students from underrepresented groups to obtain a MS degree in a very popular field that will enhance both the financial aspects and success of their future careers.

In summary, I strongly support this proposal and I believe that this new program will be very popular and attract numerous high-quality students. Please do not hesitate to contact me should you need further information.

Sincerely,

A handwritten signature in blue ink, appearing to read "J. Claar", written in a cursive style.

Jennifer Claar, PhD  
DPSS Managing Director  
Executive Office--Administration  
County of Riverside



***Council for the Advancement of Black Engineers***

3895 Brockton Ave, Riverside, CA 92501

Phone: (951) 235-8253

December 1, 2021

Re: Letter of Collaboration

Dear Prof. Tsotras,

On behalf of the Board of Directors for the Council for the Advancement of Black Engineers (CABE), I am writing in strong support of the M.S. in Computational Data Science proposed by the Department of Computer Science and Engineering and the Department of Electrical and Computer Engineering. We believe this proposed program will address current shortage in industry for students with specialized skills in Machine Learning, Programming, and Big Data Management. The proposed program will produce highly skilled students trained in both the theory and practical application sets required in field of Data Science. The proposed program has outlined a comprehensive diversity initiative to increase the number of underrepresented minorities in Data Science and computing in general. UCR is a known for its commitment to a diverse student population and the outlined initiatives will further address the commitment to increasing the diversity in the computing field, a field that generally lacks this in Academia and Industry. The M.S. Bridge program is another initiative that will ensure that students in non-computing fields will have a pathway to obtaining a M.S. in Data Science. These initiatives will prove to be models to other institutions deploying similar programs.

We strongly support the proposed M.S. in Computational Data Science and are excited about the opportunities it will create for students and professionals in the Inland Empire. We look forward to working with Dr. Tsotras on this important endeavor. Please feel free to contact me on behalf of the CABE Board if you have any questions or concerns.

Sincerely,

Dr. Brandon Davis, PhD.





CABE President

*Brandon N Davis*

Cc: CABE Board of Directors:  
Dr. Brandon Davis - President  
Mr. Elmer Thomas – Vice-President  
Dr. Brittney McKenzie - Secretary  
Mrs. Keilani Connor - Treasurer  
Mr. Keith Person, PE., - Mentorship Chair/NSBE Liaison



**Foothill Transit**

December 3, 2021

Email: [tras@cs.ucr.edu](mailto:tras@cs.ucr.edu)

Vassilis J. Tsotras  
Department of Computer Science and Engineering  
Bourns College of Engineering  
University of California Riverside  
Riverside, CA 92521

Re: **Master of Science Degree in Computational Data Science**

Dear Dr. Tsotras,

Foothill Transit provides community-oriented, environmentally friendly bus service throughout Southern California's San Gabriel and Pomona Valleys. Foothill Transit has a need to hire team members trained in both the theory and application of Data Science. We have reviewed the proposal to create a MS in Computational Data Science at the University of California, Riverside and believe that it fills an important educational requirement for the transportation industry.

Foothill Transit is applying data science to find innovative solutions to problems and questions surrounding the Inland Empire transportation system. The proposed MS degree is unique as it focuses on the Computational skills-sets that are required to store, manage, visualize, and analyze large unstructured and messy data sets. The proposed program includes a balance of courses from the two participating Departments, a capstone project and a set of electives. The capstone project is particularly of interest as it will create opportunities to connect with students on real data science projects. Foothill Transit is currently supporting the Leonard Transportation Center Research Group Challenge which draws teams of students from CSU San Bernardino, Cal Poly Pomona and UC Riverside. The proposed program and particularly the capstone project will create further opportunities of collaboration with students and faculty.

Moreover, the proposed program is a great opportunity for students from underrepresented groups to obtain a MS degree in an in-demand field. It was refreshing to note that the proposed program has outlined efforts to attract women and underrepresented minorities into Data Science. Diversifying the field of

100 S. Vincent Ave., Suite 200 • West Covina, CA 91790 W [foothilltransit.org](http://foothilltransit.org) P 626.931.7300 F 626.915.1143

**MEMBER CITIES** Arcadia, Azusa, Baldwin Park, Bradbury, Claremont, Covina, Diamond Bar, Duarte, El Monte, Glendora, Industry, Irwindale, La Puente, La Verne, Monrovia, Pasadena, Pomona, San Dimas, South El Monte, Temple City, Walnut, West Covina and Los Angeles County **A PUBLIC AGENCY**

Data Science is critical as a diverse workforce will ensure productivity and that proposed problems and solutions consider equity as part of the application or service offered.

We are particularly excited about this new program, and I believe it will produce high-quality students that will be in great demand. Thank you for your consideration. Please do not hesitate to contact me at [dbarnes@foothilltransit.org](mailto:dbarnes@foothilltransit.org) should you need further information.

Sincerely,

A handwritten signature in blue ink, appearing to read 'DJ Barnes', with a stylized flourish at the end.

Doran J. Barnes  
Chief Executive Officer





TO:  
Vassilis J. Tsotras  
Department of Computer Science and Engineering  
Bourns College of Engineering  
University of California - Riverside  
Riverside, CA 92521

December 1, 2021

Dear Professor Tsotras:

NASA Jet Propulsion Laboratory (JPL) is a federally funded research and development center and NASA field center located in Southern California. JPL has hosted many UCR interns over the past years and is proud to employ UCR graduates in various groups. JPL obviously deals with large amounts of unstructured and complex data that requires specialized skills to prepare, clean, visualize and analyze such datasets. I have worked on several machine learning projects and there is an abundance of projects that require computational data science skill sets, including visualization, pattern classification, and anomaly detection.

I am writing to express my strong support for the proposal presented to create an MS in Computational Data Science at the University of California, Riverside. This program addresses the current needs in industry for students with specialized training in Data Science, which encompasses AI, Machine Learning, large-scale software development, and big data management. The proposed curriculum is well-thought-out and balanced in terms of its focus on theory and application. It provides sufficient training in the form of the two introductory courses and offers practical training as part of the capstone project. The five elective courses allow students to focus on specific subjects based on their interest and background.

In summary, I strongly support this proposal and I believe that this new program will produce well-prepared and in-demand students soon after its launch. Please do not hesitate to contact me should you need further information.

Sincerely,

A handwritten signature in black ink, which appears to read "Zaid Towfic". The signature is stylized with a large, sweeping "Z" and "T".

Zaid J. Towfic, PhD  
Flight Communications Systems  
TEL: 818-354-1461  
zaid.j.towfic@jpl.nasa.gov



Esri Inc.  
380 New York Street  
Redlands, CA 92373  
September 16, 2020

Dear Professor Tsotras:

Esri, the worldwide leader in geographic information systems software, and an enthusiastic supporter of higher education, is pleased to endorse the University of California at Riverside's proposal to build the M.S. in Computational Data Science program.

As one of the leading technology employers in the region, Esri understands the growing need for college graduates who are prepared to contribute to innovative solutions and better products. Data Science is an emerging field and is integral to the success of Esri's solutions, which incorporate data science tools like GeoAI, ArcGIS StoryMaps and spatio-temporal analysis (to name just a few) in order to help our users achieve their goals. This new program's focus on computational aspects like Machine Learning, Artificial Intelligence, data mining and big data align particularly well with Esri's technology and future workforce needs. The program's commitment to increase diversity and the number of minority students also aligns with Esri's own values. The efforts of UCR will help ensure a supply of qualified candidates for Esri itself, as well as our partners and users. This is vitally important, as Esri and the Inland Empire as a whole seek to remain competitive in this field.

For all of these reasons, we at Esri support the establishment of the M.S. in Computational Data Science program at UCR and look forward to observing the positive impact of your graduates.

Sincerely,

A handwritten signature in cursive script that reads "Geri Miller".

Geri Miller  
Education Sector Director  
Global Business Development  
Esri

January 29, 2022

**TO:** Jason Stajich, Chair  
Riverside Division of the Academic Senate

**FROM:** Victor G. J. Rodgers, Chair  
BCOE Executive Committee

**RE:** Computational Data Science MS Proposal

---

The BCOE Executive Committee appreciates the opportunity to review the revised Computational Data Science MS Program. The revised proposal addressed all of the concerns raised in the previous submission. The committee voted unanimously in favor of the proposal.



7 April 2022

To: Jason Stajich, Chair  
Riverside Division

From: Theodore Garland, Jr., Chair, Executive Committee  
College of Natural and Agricultural Science

Re: Campus Review New Master Degree Proposal: 2nd Round-Proposal  
for a Master of Science Degree in Computational Data Science

The CNAS Executive Committee does NOT support this proposal: not a single member voted in favor.

We were unimpressed by the revisions.

Changing the title was the main thing noted. Other edits were non-substantive.

The section on diversity seems boilerplate and needs to be more detailed/specific.

We worry about the unintended consequences of duplication of learning goals and materials across the colleges.

We understand that the current budget model does not encourage cross-college collaborations, especially when a revenue-generating program is being developed. This is a structural problem with the budget model, and needs to be brought to the attention of the Provost and Chancellor.

Cheers,

A handwritten signature in black ink that reads "Ted Garland".

**COMMITTEE ON DIVERSITY, EQUITY, & INCLUSION**

April 15, 2022

To: Jason Stajich  
Riverside Division Academic Senate

From: Katherine Stavropoulos, Chair  
Committee on Diversity, Equity, & Inclusion

Re: [Campus Review] 2nd Round-Proposal for a Master of Science Degree in  
Computational Data Science

The DEI committee reviewed the 2nd Round-Proposal for a Master of Science Degree in Computational Data Science document, which addresses issues of diversity previously raised by the committee. However, it remains unclear how the proposers will undertake a quantitative diversity assessment through anonymous student surveys. Depending on the projected size of the program, asking students to provide information about their gender/race/ethnicity may undermine efforts to anonymize the survey and could lead to either: a) students not feeling comfortable disclosing concerns, or b) a lack of participation from students. It would be helpful to have more information on how the proposers plan to complete these surveys (e.g. what questions do they plan to ask, how will they promote transparency/participation from students, etc).

Furthermore, the list of organizations meant to increase the participation of underrepresented groups in data science is great. It would be even better if the proposed MS program made it a goal to contribute to these programs or advertise within them. The grant mentioned might already be a part of that, but it isn't very clear from what is written.





## *Academic Senate*

### **GRADUATE COUNCIL**

April 21, 2022

To: Jason Stajich, Chair  
Riverside Division

From: Don Collins, Chair  
Graduate Council

Re: [Campus Review] New Master Degree Proposal: 2nd Round-Proposal for a  
Master of Science Degree in Computational Data Science

Graduate Council reviewed the revised proposal for a Master of Science Degree in Computational Data Science at their April 21, 2022 meeting and approved this version of the proposal.



## *Academic Senate*

### **PLANNING & BUDGET**

April 12, 2022

To: Jason Stajich, Chair  
Riverside Division

From: Dana Simmons, Chair  
Committee on Planning and Budget

**RE: [Campus Review] New Master's Degree Proposal: 2nd Round-Proposal for a Master of Science Degree in Computational Data Science**

Planning & Budget (P&B) reviewed the revised proposal for a new Master of Science degree in Computational Data Science at their April 12, 2022 meeting. P&B supported this revised version of the proposal that now includes a letter of support from BCOE Dean Lynch.

Professor Jason Stajich, Chair  
Riverside Division

May 3rd, 2022

Dear Prof. Stajich,

Please find below our response to the comments from the Committee on Diversity, Equity, and Inclusion and the CNAS Executive Committee on our proposal for a MS in Computational Data Science. We would like to thank the committee members for their careful consideration of our proposal.

Sincerely,



V. Tsotras

on behalf of the MS in Computational Data Science Program Committee

## **Response to the Committee on Diversity, Equity, and Inclusion**

In summary, the CODEI committee: (1) asked for ways to ensure that student survey data would remain anonymous/not be compromised by the demographic information collected in surveys, and, (2) suggested partnering with organizations related to underrepresented groups in data science.

We would like to clarify that the demographic data will not be collected by our surveys. Instead we will use data collected by Institutional Research (together with their application, admission and retention data). Our surveys will be optional and will focus on assessing the student's self-efficacy in various topics / technical tools and whether they perceive they have sufficient opportunities as part of the program (opportunities can include mentorship, internships, summer research, etc.) To create and administer the surveys we will use an outside company and such surveys are approved by IRB. As the company holds the survey responses we do not have access to who responded and how. Because the demographic data will be collected in a separate survey from the educational component, there will not be a way of cross-correlating the information.

With respect to suggestion item (2), we note that we have already established collaborations/partnerships with the URM organizations mentioned in the proposal through existing research projects/grants. We will use these partnerships to advertise the program and attract URM students. Examples include local HSIs (like CSUSB, CalPoly-Pomona, California Baptist University, La Sierra University and all local community colleges), the Council for the Advancement of Black Engineers (CABE), the Computing Alliance for Hispanic Serving Institutions (CAHSI) network as well as the local WIDS conference (Women in Data Science).

## Response to the CNAS Executive Committee

The CNAS EC response raised four points:

1. The committee noted the changed title. This title change (Computational Data Science) was a direct response to previous concerns that the program would block other similar efforts. We made the degree title more specific to allow for other programs to concentrate on other aspects of Data Science. We are not against such programs and have specifically changed the title of this proposed program to make it clear that such programs could exist. We have also made it clear to the Statistics Department that we would not stand in the way of any analytics-oriented Data Science Masters (and would help in its creation if requested).
2. CNAS EC stated that the diversity section needed to be more specific. The reviewed proposal had 13 pages of details that were developed specifically for this program. We have addressed CODEI's remaining minor points above.
3. The response states a "worry about the unintended consequences of duplication of learning goals and materials across the colleges." The courses for this program have already been approved. If this worry relates to duplication of program-level goals, its focus on the computational aspects delineates it from other potential programs (as noted in point 1). Other broad disciplines have specialized graduate programs, focusing on specific areas, as this proposal does. Our pedagogical goal is to provide students who have a computational undergraduate degree a Masters program that focuses on the computational aspects of Data Science, including data storage, indexing, and collection, and the algorithmic considerations for its processing. With a 4-quarter program, there is not time for the breadth necessary to cover all aspects of Data Science.
4. The response raised general concerns about revenue sharing. These are not points that can be solved or addressed by this proposal. Nor should such concerns restrict the ability of a proposal to move forward.

Dear Prof. Stajich,

Thank you for the comments on the MS Data Science program proposed by the CSE and ECE departments. We have carefully reviewed the concerns that were expressed and believe we have addressed each of them in this document. One of the more significant concerns was the reduced role of the Statistics Department in the MS program relative to their role in the undergraduate Data Science program. You will find that we have paid particular attention to this concern and believe we have clarified the pedagogical reasons for designing the MS Data Science degree in this particular way. To avoid confusion with statistical data science and mixed statistical / computational data science programs, we have changed the name to “MS in Computational Data Science”.

Further, we reached out again to the Statistics Department and have already committed to help Statistics in creating a Data Science MS track focusing on the Analytical part of Data Science, when the Department is ready to explore that option. The Chair of the Statistics Department has informed me that the Statistics faculty is now on board with the proposed new program.

Below we address in detail the various comments we received from the Senate committees. The main concerns were with: (1) the diversity plan, (2) the reduced role of the Statistics department, and, (3) clarification of the BCOE support for the resources that will be needed by the new program.

### **Diversity plan:**

We appreciate the feedback that made clear we did not adequately articulate our plan for supporting diversity in the program. We fully rewrote the related section, clarifying our vision and our plan to realize our vision. This includes enumeration of the various efforts that will be integrated into our management for the program that include following a set of best practices to improve diversity suggested in a recent report by the National Academies of Sciences, Engineering, and Medicine<sup>1</sup>.

### **Reduced role of the Statistics Department relative to the BS Data Science program:**

We attempted to be transparent about our approach. Much of what we articulated in this section about focusing on computational data science was discussed with some members of the Statistics Department (including the Chair) before we began the detailed development of the

---

<sup>1</sup> National Academies of Sciences, Engineering, and Medicine. 2018. Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments. Washington, DC: The National Academies Press. <https://doi.org/10.17226/24926>.

proposal. As tends to happen with communication in a large organization, we clearly did not succeed in communicating our plans to all stakeholders.

In the proposed MS in Computational Data Science, we focus only on the computational aspects of Data Science. This is clarified by the updated title of the program. By computational we mean the systems (hardware, algorithms, software) that provide for the storage, indexing, retrieval and understanding of the data at the volumes and speeds necessary for computational analysis. The updated name also allows the Statistics Department to create a Data Science MS focusing on the statistical part of Data Science if they decide so in the future.

We understand that our decision to focus on computational data science in the proposed MS degree program raised concerns, particularly in that this could close the door to the Statistics Department either participating in future MS Data Science degree programs with BCOE, or potentially even hinder them launching their own program. This was not the intent. The decision to focus on computational data science was made based on the anticipated background of the students that the program was designed to serve. Our target is students with a computational undergraduate background. The goal is for them to complete the MS Computational Data Science degree in 4-5 quarters, similar to related MS programs like computer science and computer engineering. Few, if any, of the students we believe will be attracted to this program will have an undergraduate degree in Statistics. This would result in their having to take a number of undergraduate statistics courses to prepare them to take the graduate statistics courses. The result would be an increased time to degree if we were to require both computational data science and statistics based data science courses. As mentioned, we reached out to again to the Statistics Department and clarified our intent and willingness to support them in the creation of a MS in Analytical Data Science when Statistics wants to explore that path.

#### **Resources needed by the new program:**

The Bourns College of Engineering, in the attached BCOE Dean's letter, has made a clear commitment to provide the resources that will be required to support this program.

We would like to take this opportunity and thank all the involved senate committees and their members for their reviewing and commenting on our proposal and for their constructive recommendations that have enabled us to strengthen this proposal.



Sincerely,  
Vassilis Tsotras  
on behalf of the MS in Computational Data Science Program Committee

## **Response to the comments by the Committee on Diversity, Equity, & Inclusion:**

Thank you for your recommendations. These have enabled us to strengthen our articulation of our commitment to DEI. We have revised Section 1.5 (Contributions to Diversity) to include our diversity vision and our plan to realize our vision. As requested, we have added current UCR and national data on women and URM enrollment in related programs (Computer Science/Electrical Engineering). Our goal is to achieve the BCOE average enrollments for women (22.8%) and URMs (12.6%) by the 2nd year of the program and then gradually increase their participation to 30% and 15% (respectively) by year 4, aiming at 40% and 20% (respectively) by year 5.

To reach these goals, we will follow a multipronged approach that addresses diversity at different levels in: recruitment, curriculum and pedagogy, outreach, assessment, and faculty. In each of these areas we will follow the set of best-practices for improving diversity in computing discussed in a recent report by the National Academy of Sciences<sup>2</sup>. The two departments participating in the new MS in Computational Data Science are fully committed to Broadening Participation in Computing (BPC) and have created departmental BPC plans to that effect. Various of the BPC initiatives have been made possible by a number of recent grants, as discussed in the proposal, that have been awarded to faculty in the program and/or to BCOE with focus on improving diversity in computing.

There was a concern expressed about partnering with the Naval Surface Warfare Center in Corona (NSWCC), "...in a way that doesn't take into account DEI issues". We would like to clarify that the proposed program is not affiliated in any way with, nor was its design affected by NSWCC. We mentioned NSWCC in the proposal as one of the local employers (the largest in the Inland Empire in this field) that is very much interested in employing more people with Data Science related expertise. We further note that Data Ethics is a very important issue addressed in two parts of the proposed curriculum: CS252/EE251A (Data Analytics and Exploration) and CS/EE279 (Capstone)<sup>3</sup>. We do anticipate that our students' exposure to Data ethics issues will be beneficial to NSWCC and other employers who hire them. NSWCC is far from the only local employer interested in this program. We have included letters of support from other local employers (government and businesses) that are interested in hiring the graduates of the proposed program.

---

<sup>2</sup> National Academies of Sciences, Engineering, and Medicine. 2018. Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments. Washington, DC: The National Academies Press. <https://doi.org/10.17226/24926>.

<sup>3</sup> In addition, there is a current effort to create a new Data Ethics course for the Data Science Major, as well as a graduate version of that course which we will consider after its (expected) approval.

**Response to the comments by the Graduate Council Committee:**

We understand the concerns expressed. We hope that you will find the proposed program changes have addressed those concerns.

1. *Recommendation to involve Statistics.*

We agree that with the original name of the program there would have been confusion and that this could have hindered the development of future data science programs by other colleges. We believe we have addressed this issue with the name change and with the explanation of the reason requiring graduate statistics courses would lead to an increased time to degree.

There are a plethora of MS programs related to Data Science, some of them hosted in Statistics Departments, some in Computer Science, some in Business and some in combinations of these; we chose to host the new MS in the Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE) departments based on the educational needs of the students we are targeting. In the limited duration of a MS degree there is not enough time to cover all aspects of Data Science. As examples of computational data science programs we note CMU (which is among the top universities in AI research) has a program in Computational Data Science offered by their School of Computer Science. Similarly, Temple University has an MS in Computational Data Science program offered by their Computer and Information Science Department. Within the UC, the most similar is the MS in Machine Learning and Data Science offered by the Department of Electrical and Computer Engineering at UCSD. We feel that changing the name to Computational Data Science clarifies the focus of the proposed program.

The focus on Computational Data Science is supported by: (1) the two departments (CSE and ECE) that have strong expertise and have already been offering most of the courses needed by this focus; (2) we aim to attract students that do not have broad statistical background, thus asking them to attend graduate core courses in statistics would necessitate an increased time to graduation, significantly more than what is taken by other MS programs in BCOE and in engineering colleges more broadly<sup>4</sup>; (3) there are many computational data science jobs that focus on AI, Machine Learning, Data Mining, Big Data, etc. By computational we mean the systems (hardware, algorithms, software) that provide for the storage, indexing, retrieval and understanding of the data at the volumes and speeds necessary for computational analysis.

We have changed the original name of the proposed MS program from 'Data Science' to 'Computational Data Science' so as to better clarify the focus of the program. This name change leaves open the possibility for Statistics to create another MS that focuses on the statistical aspects of Data Science if they decide so in the future.

The proposed program would be the second MS at UCR related to Data Science. Recently, the MS in Business Analytics program was approved and it is run by the School of Business and the Statistics Department. Business Analytics is the intersection of Business and Data Science. The two units decided that this collaboration is the best for the students that the new program will

---

<sup>4</sup> We expect that other related graduate courses (Statistics included) will be included as electives in the proposed program in the near future.



serve and did not include BCOE. Many other Business Analytics programs (NYU Stern as an example) use a strong coverage of big data and databases (both CSE topics) in their program. This does not mean that every Business Analytics program should include CSE courses; rather this choice should be guided by the needs of the students the program will serve.

To be clear, we very much value the collaboration with the Statistics Department on the Data Science BS Major. This new undergraduate major was initiated by CSE faculty who invited Statistics faculty to collaborate in the design of the major since this was the appropriate educational decision for that 4-year program. We note that we discussed our plans to develop a MS Data Science program with members of the Statistics Department before we submitted the original MS proposal and explained the reasons for the choice. The concerns raised with respect to our original proposal indicate that this discussion was not broad enough. We have met with the Statistics Chair before completing this revised MS submission to: (i) discuss the proposed name change that will allow Statistics to create a new program, and (ii) consider possible Statistics courses that we could include as electives in the proposed computational data science program. We have also committed to help Statistics in creating a Data Science MS track focusing on the Analytical part of Data Science, when the Statistics Department is ready to explore that option. We also agreed that such elective courses can be added in the future. The Chair of the Statistics Department has informed us that the Statistics faculty is now on board with the proposed new program.

## *2. Diversity concerns.*

We have expanded the “contribution to diversity” section to include a vision statement and a detailed plan. Please see similar concerns and our reply to the DEI committee.

With respect to the assessment mechanism for diversity, we note that both participating departments have implemented such a mechanism as part of the data collection for their Broadening Participation Committees (BPC). Similarly, such diversity assessment is required as part of a new NSF grant to create Data Science pathways for Inland Empire students. Data will be collected with the help of Institutional Research to evaluate admission, demographics and retention.

As for faculty diversity, the participating program faculty of the proposed MS in CDS includes 5 women and two URMs (1 Pacific Island and 1 Latinx). Even though the proposed program is not a department that can hire its own faculty, there are currently three faculty searches in the participating departments (CSE and ECE) in areas related to Data Science. Special efforts will be taken for these and future searches to (i) increase the representation of women and minorities in the pool of candidates by announcing the recruitments in related forums including WiML, BlackInAI, LatinXinAI, QueerInAI, HSI institutions and the National Society of Black Engineers and (ii) include women/minorities on the search committees in an effort to avoid unconscious bias.

Note that we mention the military industry in Corona, as just an example of a large local employer. Other local industries (e.g. ESRI) as well as throughout California will benefit by recruiting our graduates.

*3. Comments from other UCs about the proposed program.*

Before the first submission we contacted various other related UC programs but we did not get a response before that submission. After the one month period passed we proceeded with the submission so as to not delay the program further. Meanwhile we have asked and received such letters that are included in the second submission (this includes: UCD, UCI, UCLA, UCSB, UCSC and UCSD; we also have letters from NYU and CMU).

**Response to the comments by the Committee on Planning and Budget:**

The support for the resources required by the new program is now discussed in the BCOE Dean's letter.

**Response to the comments by the CNAS Executive Committee:**

We thank you for the comments on the original proposal. We hope that you will find that the proposed program changes have addressed those concerns.

1. *"In the larger scheme of things, Data Science should be applied. If approved in Data Science in BCOE, then no one else can have Data Science Masters."*

We definitely understand this concern about the new program's name. Our purpose was not to block others (and this was clearly indicated to the Statistics Chair when we informed the chair about the plan to create a new MS in Data Science). We have changed the name of the proposed program to MS in Computational Data Science. This name better describes the focus of the proposed program on the foundational computing principles of data science. It also leaves open the possibility for Statistics to create another MS focusing on the statistical side of Data Science or to participate with BCOE in developing such a program.

2. *"The BS is joint STAT and Computer Engineering. But here Computer Engineering decided to launch the Masters on their own. We understand that STAT feels unfortunate to be left out of this. Data science should be a marriage between STAT and Computer Engineering, especially since they are doing the BS together. If we're looking at the viability of the program on the whole, it's missing STATs courses."*

Before we started the design of the original MS in Data Science proposal we informed the Chair of Statistics and the co-director of the BS in Data Science about the plan and why we think that this new MS should be between CSE and ECE. In particular, we explained that the time restrictions that an MS program introduces places limits on the material that can be covered; this is quite different from the 4-year period of a BS. At the time, the Statistics Chair indicated that Statistics had just put together a (data science related) program with the Business School (MS in Business Analytics) and suggested looking at it to see if it might be helpful as an example of creating a DS program with BCOE. We looked at the Statistics courses offered as part of that new Business Analytics program, namely: Statistics for Business Analytics (STAT232), Statistical Data Mining (STAT208), Discrete Data Analysis (STAT205) and Statistical Computing (STAT206). Among them, STAT232 is Business focused, while STAT 208 deals with data mining topics from a statistical perspective. In BCOE, we have courses which cover related topics but from a computational and algorithmic perspective, which is the focus of the MS CDS program. We examined STAT205 and STAT206 but they have a different focus on foundational statistical methods.

The MS program is close to a professional degree program that prepares students for jobs. There are many data science jobs that are based on the computational aspects of Data Science, i.e. focusing on AI, Machine Learning, Data Mining, Big Data, etc. By computational we mean the systems (hardware, algorithms, software) that provide for the storage, indexing, retrieval and understanding of the data at the volumes and speeds necessary for computational analysis. The proposed MS in Computational Data Science aims to prepare our students for such computational focused jobs; most of these jobs do not require a strong statistical

background, and the ones that do would likely recruit from a statistics oriented data science program. Nevertheless, we are open to adding approved Statistics courses related to Data Science as electives; we have recently reached out to the Statistics Chair for such course suggestions. We expect that more data science courses (not only from Statistics) can be added as electives in the future.

Having a common BS program between CSE and Statistics does not necessarily mean that all other data science related programs should be in common. An example is the MS in Business Analytics. Business Analytics is the intersection of Business and Data Science. Moreover, CSE has a related undergraduate BS, the Computer Science with Business Applications (CSBA) major where students take basic CS and Business courses. Nevertheless, the Business School and the Statistics Department designed the new MS in Business Analytics based on the educational needs of the students they target and BCOE respects this decision. Similarly, the MS in Computational Data Science is designed for students with a STEM background that are interested in the computational aspects of Data Science.

Further, we reached out again to the Statistics Department and have already committed to help Statistics in creating a Data Science MS track focusing on the Analytical part of Data Science, when the Department is ready to explore that option. The Chair of the Statistics Department has informed us that the Statistics faculty is now on board with the proposed new program.

2. *“The proposal notes that women are extremely underrepresented. They say the structure of the program will include a more diverse group of students, but no strategy for making it happen is included.”*

We have expanded the “contribution to diversity” section to include a vision statement and a plan. To reach these goals, we will follow a multipronged approach that addresses diversity at different levels, in: recruitment, curriculum and pedagogy, outreach, assessment and faculty. In each of these areas, we follow a set of best-practices for improving diversity in computing, discussed in a recent report by the National Academy of Sciences<sup>5</sup>. Moreover, the two departments participating in the new MS in Computational Data Science are fully committed to Broadening Participation in Computing (BPC) and have created departmental BPC plans to that effect. Various BPC initiatives have been made possible by a number of recent grants (discussed in the proposal) that have been awarded to faculty in the program and/or BCOE with focus on improving diversity in computing.

3. *“We expect students to come into the MS in Data Science through the traditional engineering pathway, but maybe also people with biology degrees will turn up. It seems that having a broader base program would make that more likely to happen (to diversify the program). What about capstone courses that cross colleges and apply/execute the data science. Some felt that this is what industries are looking for.”*

---

<sup>5</sup> National Academies of Sciences, Engineering, and Medicine. 2018. Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments. Washington, DC: The National Academies Press. <https://doi.org/10.17226/24926>.

While we focus on STEM students with computational background, we have created a bridge program that will offer courses so that students with non-CS degrees can cover key undergraduate material and then proceed to pursue an M.S. in Computer Science, Computer Engineering, or Computational Data Science. We expect that this path will be attractive to students from fields like Biology/Neuroscience, Sociology, Economics, Earth Sciences etc.

The capstone course focuses on the design and execution of a large-scale data science capstone project that has a focus on real-world applications. We will be collaborating with local industry (ESRI etc.) to identify such real-world problems. We will also reach out to faculty in other departments to identify possible capstone projects.

4. *"In the computer science proposal, it is argued that UCSD, UCLA, and UCB all have data science programs through engineering school. However, we would like to point out that (1) those programs are all online MS programs, (2) BCOE already has an online MS program in data science, and (3) the Department of Statistics is already teaching two online courses in BCOE's the online MS program. In addition, the following are examples of data science programs in other universities that are joint programs, which creates a stronger program:"* (...links provided for the UCI, Harvard, Duke, Columbia, Uof Washington and Stanford Data Science MS programs)

In the previous proposal we only provided information about known programs at various UCs and which college offers them. We also mentioned that the majority of these programs are on-line or professional (self-supported). We never argued that UCSD, UCLA and UCB all have data science programs through engineering. Some of these programs are offered by Engineering schools, others by School of Information, others by School of Business, etc. At the time of the first submission, UCI had only a MS in Business Analytics. In Fall 2021 UCI started a professional MS program in Data Science offered by their Bren School of Information and Computer Science (that houses also the Department of Statistics).

BCOE already has an online MS program that offers a Data Science concentration. The Department of Statistics teaches two online courses in that program. However, these courses are not core courses. Instead, the MSOL program has 5 core courses (ENGR200-205) that focus on engineering management, systems engineering, technology innovation etc. There is also a long list of electives including CSE, EE, Physics and Statistics courses from which students can take 4 courses. Our proposed MS in Computational Science is very different as it focuses on building the computational aspects of data science. As mentioned earlier, we have approached the Statistics department to suggest graduate courses that can be included as electives in the proposed MS program.

There are a plethora of MS programs related to Data Science (including the ones provided by the committee), some of them hosted in Statistics Departments, some in Computer Science, some in Business and some in combinations of these. At UCR the only other Data Science related program is the new MS in Business Analytics between the Business School and

Statistics. We chose to host the new MS in the Computer Science and Electrical Engineering departments for educational purposes. In the limited duration of a MS degree there is not enough time to cover all aspects of Data Science. MS programs focusing on the Computational part of Data Science are not unique. Similar programs appear elsewhere, for example CMU (which is among the top universities in AI/ML research) has a program in Computational Data Science offered by their School of Computer Science; and Temple University has an MS in Computational Data Science program offered by their Computer and Information Science Department. Within the UC, the most similar is the MS in Machine Learning and Data Science offered by the Department of Electrical and Computer Engineering at UCSD. We feel though that the name Computational Data Science better describes our proposed program.




*Academic Senate*

**EXECUTIVE COUNCIL**

January 13, 2021

**To:** Prof. Vassilis Tsotras  
Department of Computer Science & Engineering

**From:** Jason Stajich, Chair   
Riverside Division

**RE:** Proposal for a Master of Science Degree in Data Science

At its January 11, 2021 meeting, Executive Council reviewed the proposal for a Master of Science Degree in Data Science.

The reviewing committees had several concerns and questions. I attach responses from the Committee on Diversity, Equity, & Inclusion, Graduate Council, Committee on Planning & Budget, BCOE College Executive Committee, and CNAS Executive Committee.

Thanks,  
Jason

Cc: Cherysa Cortez, Executive Director, Academic Senate





**COMMITTEE ON DIVERSITY, EQUITY, & INCLUSION**

November 16, 2020

To: Jason Stajich, Chair  
Riverside Division Academic Senate

From: Xuan Liu, Chair  
Committee on Diversity, Equity, & Inclusion

Re: New Master Degree Proposal: M.S. Degree in Data Science

The Committee on Diversity, Equity, and Inclusion (CoDEI) reviewed the proposal M.S. Degree in Data Science. The Committee notes the proposed DEI plan only discussed current gender of students enrolled but not underrepresented minorities nor how they will actively recruit.

Per the Format for the Graduate Degree Program Proposal outlined in Appendix B of the Coordinating Committee on Graduate Affairs (CCGA) Handbook, August 2019 revision, "All proposals must include (a) a vision for how the program will advance UC's goals for diversity and (b) a plan that details what steps the program will take in its first five years to move it toward the identification, recruitment, and retention of underrepresented minority students and faculty. The proposal should clearly document the ways in which the program will evaluate its diversity goals."

In addition, some members also expressed concern on the proposed partnering with the Naval Warfare Center in a way that doesn't take into account DEI issues which is potentially relevant, whether because the US military is increasingly attracting underrepresented students or in how critical thinking and diverse sets of students would be beneficial for them. Those members are concerned that some students may protest against an affiliation particularly in this moment of time where policing and surveillance are under heightened surveillance because of the violence toward Black and Brown people.



## GRADUATE COUNCIL

## *Academic Senate*

December 18, 2020

To: Jason Stajich, Chair  
Riverside Division

From: Amanda Lucia, Chair  
Graduate Council

A handwritten signature in blue ink, appearing to read "Amanda Lucia".

Re: Proposal for a Master of Science in Data Science

Graduate Council reviewed the proposal for an MS degree in Data Science at their December 10, 2020 meeting.

The Council would recommend the involvement of the Statistics department and its faculty in the proposal to strengthen the program considering the fact that the majority of national data science programs (if not all) contain statistics, and the BS in Data Science and the Online Master's Degree with a specialization in Data Science at UCR both involve statistics. Some of the reputed data science programs can be found at UCI, Harvard, Stanford, Duke, Columbia, and the University of Washington.

The Council also had concerns with the diversity component required in new program proposals. The contribution to diversity in the proposal was rather vague and it was not clear how the program will address diversity. The Coordinating Committee on Graduate Affairs (CCGA) will require an assessment mechanism for the diversity of the program be in place, the Council is recommending this be added to the program proposal to alleviate delay of approval at the systemwide level. The Council noted the lack of gender representation and recommends a mechanism for addressing faculty diversity also be added. The proposal tends to focus on the recruitment of students, faculty of color, and white women faculty as a way to address gender and race-based diversity. But there is no concrete description of specific steps taken in the past or planned for the future. This looks to be a weakness of the program and, if it is, it should be confronted head-on. The proposal mentions in several places the pedagogical and research/scholarly goals of providing "real life solutions" and addressing "real world problems." How might this goal be intertwined with that of genuine diversification and desegregation as a problem not merely of bodies (students of color, faculty of color, white women faculty) but a body of


knowledge? For example, the cutting-edge work on race and algorithms by Black studies scholars like Safiya Umoja Noble (UCLA). There are a couple of ways to address this already embedded in the proposal: (a) Draw on the sample syllabus appended to the proposal - the week on ethics and bias in information science; (b) Another approach would be in the description of a potential bridge for students who do not have undergraduate degrees in the quantitative sciences but who could qualify by taking certain courses. What is the importance and benefit of recruiting students with a non-quantitative background? (c) Yet another approach would be to amplify and think more creatively about UCR relations with its surrounding communities and business landscape. The proposal focuses on the military industry in the inland empire but there must be other industries and phenomena that UCR is strategically poised to study. Many UCR students work part-time as warehouse workers and their lives are intricately intertwined with what the geographer Juan De Lara calls the “black box of globalization,” e.g. logistics and supply-chain of consumption and distribution.

The proposal did not include comments from other UCs about the proposed program. It may be worth following up with other UC campuses (for example, UCI has a similar data science program), so that their feedback can be included in the proposal.



December 15, 2020


To: Jason Stajich, Chair  
Riverside Division

From: Katherine Kinney, Chair   
Committee on Planning and Budget

**RE: [Campus Review] New Master's Degree Proposal: M.S. Degree in Data Science**

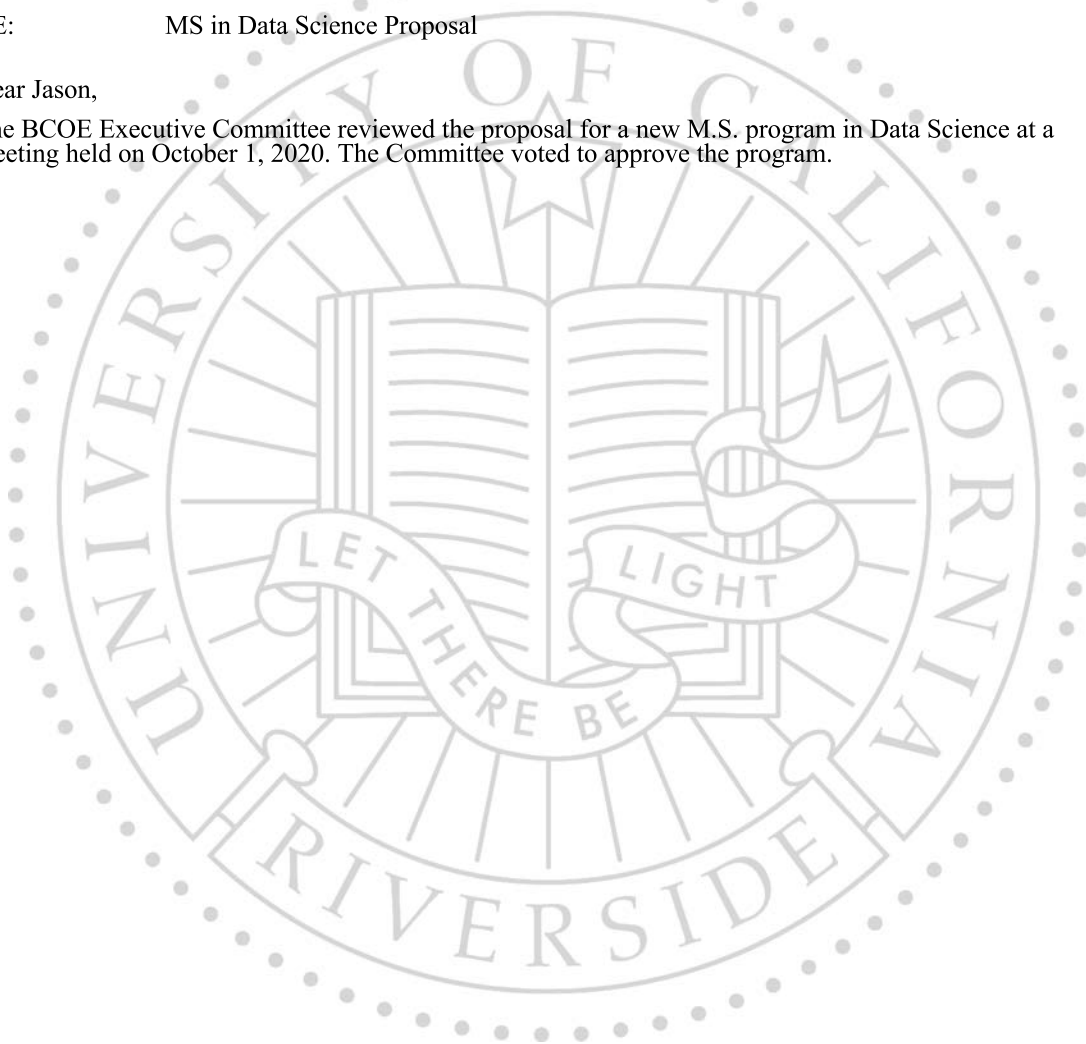
The Committee on Planning & Budget discussed the proposal for a new Master of Science degree in Data Science at their December 15, 2020 meeting. The committee noted that the program will require additional resources (office space and one FTE for administrative support); however, there is no commitment to provide funding for these resources in any of the support letters. The source of support for the additional resources needs to be defined in the proposal.

October 23, 2020

TO: Jason Stajich, Chair  
Academic Senate  
FROM: Philip Brisk, Chair   
BCOE Executive Committee  
RE: MS in Data Science Proposal

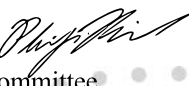
Dear Jason,

The BCOE Executive Committee reviewed the proposal for a new M.S. program in Data Science at a meeting held on October 1, 2020. The Committee voted to approve the program.



January 11, 2021

TO: Jason Stajich, Chair  
Academic Senate

FROM: Philip Brisk, Chair   
BCOE Executive Committee

RE: MS in Data Science Proposal

Dear Jason,

The CNAS Executive Committee's review of the MS Degree in Data Science proposal noted that the Statistics Department feels unfortunate to be left out of this.

I was surprised to see this comment, because my understanding was that Vassilis Tsotras, who submitted the proposal, had been in communication with the Statistics Department. I asked Vassilis if he could confirm and summarize prior communication while the proposed MS Degree was under development, to make sure that the written record is accurate. He provided a written response this morning, which I am transmitting along with this letter.



PROFESSOR VASSILIS J. TSOTRAS  
DEPARTMENT OF COMPUTER SCIENCE  
Bourns College of Engineering  
University of California  
Riverside, CA 92521  
Tel: 951-827-2888  
Fax: 951-827-4643  
Email: tsotras@cs.ucr.edu

January 11<sup>th</sup>, 2021

Prof. Philip Brisk  
Chair, BCOE Executive Committee

Re: the MS in DS proposal

Dear Philip,

Thank you for bringing this important question to my attention. The reason for this letter is to clarify that we did contact Statistics even before we started designing the MS program in DS. Our intent has never been to keep them in the dark. Indeed, it is worth noting that the BS in DS was a CSE initiative, and we reached out to Statistics to bring them on board.

On 4/3/20, I informed Prof. Xinping Cui (Statistics chair) and Prof. Yehua Li (co-director from Statistics of the BS in Data Science) by email of our plan to begin designing a new MS in Data Science within BCOE. I explained that the tight 4-quarter MS timeline required focus, and did not permit time to include extensive coursework from both the computational part (data processing, data mining, etc.) of Data Science and the analytic part (that involves the statistical methods). While there may be other programs elsewhere that provide more breadth and less depth, that is not our goal. Instead our plan is to develop an MS in DS that focuses more on the computational part of Data Science. At the BS level, in contrast, the 4-year timeline permits more time and more flexibility.

I also indicated we would have no objection to Statistics creating their own version of a MS in Data Science.

Prof. Cui replied on 4/7/20, informing us that Statistics had just put together a similar program with the Business School (MS in Business Analytics) and suggested looking at it to see if it might be helpful as an example of creating a DS program with BCOE. We looked at the Statistics courses offered as part of that new program, namely: Statistics for Business Analytics (STAT232), Statistical Data Mining (STAT208), Discrete Data Analysis (STAT205) and Statistical Computing (STAT206). Among them, STAT232 is Business focused, while STAT208 has large overlap with Data Mining courses offered in CSE and ECE. We examined further STAT205 and STAT206 but they are at a more advanced level and would require our target students to take a significant number of prerequisites in Statistics.

After you contacting me, I reached out to Prof. Cui and Prof. Li and on 1/8/21 we met over Zoom so as to clarify any misunderstandings. Prof. Shelton (from CS) also attended (Prof. Shelton is a member of the MS in DS design committee). I shared with them an earlier version of this letter as well as the submitted MS in DS proposal. I now believe that some of the misunderstandings between the CSE and Statistics colleagues may have been due to the start of the university closure (because of the pandemic) which did not facilitate easy access/discussions. During our meeting we agreed that both the CSE and Stats departments value the special relationship that the new DS undergraduate major has created. We also discussed the focus of the MS proposal as well as its name. Prof. Cui informed the Statistics faculty about what was discussed and shared the earlier version of the letter and the proposal with them. Prof. Cui emailed me on 1/10/21 with the Statistics concerns which can be summarized as: (i) an apparent lack of the interdisciplinary collaborative nature among Computer Science, Mathematics, Statistics and domain expertise in the data science proposal, and, (ii) the name of the proposed MS.

From our perspective, the strong computational focus on our proposed program, and the tight 4-quarter timeline precludes extensive coursework from Statistics or other domains. Note that the proposed program also offers a list of electives which can be enhanced with more interdisciplinary coursework. We are open to changing the name of the

proposed program to reflect its computational focus better, so that Statistics retains the option of creating their own program in Data Science down the road.

We will of course wait for the official response from the Senate and answer this and any other issues raised about the MS in DS proposal. Please let me know if I can provide any further clarifications.

Sincerely,

Sincerely,

A handwritten signature in black ink, appearing to read "V. Tsotras".

Vassilis J. Tsotras  
Professor and Co-Director of Data Science Major  
Dept. of Computer Science and Engineering  
University of California, Riverside  
tel: 951-827-2888; email: [tsotras@cs.ucr.edu](mailto:tsotras@cs.ucr.edu)





23 November 2020

To: Jason Stajich, Chair  
Riverside Division

From: Theodore Garland, Jr., Chair, Executive Committee  
College of Natural and Agricultural Science

Re: Campus Review: New Master Degree Proposal: M.S. Degree in Data Science

The CNAS Executive Committee had many comments and concerns about this proposal, including:

In the larger scheme of things, Data Science should be applied. If approved in Data Science in BCOE, then no one else can have Data Science Masters.

The BS is joint STAT and Computer Engineering. But here Computer Engineering decided to launch the Masters on their own.

We understand that STAT feels unfortunate to be left out of this. Data science should be a marriage between STAT and Computer Engineering, especially since they are doing the BS together.

If we're looking at the viability of the program on the whole, it's missing STATs courses.

The proposal notes that women are extremely underrepresented. They say the structure of the program will include a more diverse group of students, but no strategy for making it happen is included.

We expect students to come into the MS in Data Science through the traditional engineering pathway, but maybe also people with biology degrees will turn up. It seems that having a broader base program would make that more likely to happen (to diversify the program).

What about capstone courses that cross colleges and apply/execute the data science. some felt that this is what industries are looking for.

Cheers,

A handwritten signature in black ink that reads "Ted Garland".

**UC RIVERSIDE** UNIVERSITY OF CALIFORNIA | College of Natural  
& Agricultural Sciences  
Executive Committee

18 December 2020

To: Jason Stajich, Chair  
Riverside Division

From: Theodore Garland, Jr., Chair, Executive Committee  
College of Natural and Agricultural Science

Re: Additional Commentary on Campus Review: New Master Degree Proposal: M.S.  
Degree in Data Science

The Executive Committee previously commented on this proposal on November 23, 2020 and has since come into receipt of additional information supporting our initial concerns. In the computer science proposal, it is argued that UCSD, UCLA, and UCB all have data science programs through engineering school. However, we would like to point out that (1) those programs are all online MS programs, (2) BCOE already has an online MS program in data science, and (3) the Department of Statistics is already teaching two online courses in BCOE's the online MS program. In addition, the following are examples of data science programs in other universities that are joint programs, which creates a stronger program:

UCI data science MS program: <https://mds.ics.uci.edu/>

Harvard University data science MS program: <https://www.seas.harvard.edu/applied-computation/graduate-programs/masters-data-science>

Duke University data science MS program (this program is interdisciplinary): <https://datascience.duke.edu/academics>

Columbia University data science MS program: <https://datascience.columbia.edu/education/programs/m-s-in-data-science/>

The University of Washington data science MS program (this program is jointly lead by six departments): <https://www.washington.edu/datasciencemasters/curriculum/>

Stanford University data science MS program: <https://statistics.stanford.edu/academic-programs/graduate-programs/ms-statistics-data-science>

Cheers,



*A Proposal for a*

**MASTER OF SCIENCE DEGREE IN DATA SCIENCE**



Marlan and Rosemary Bourns College of Engineering

University of California – Riverside

Riverside, CA 92521

Submitted by

Vassilis Tsotras on behalf of the MS Data Science Program Committee

Professor, Department of Computer Science and Engineering

Director, Data Science Center

Program Director, Data Science Undergraduate Program

M.S. Data Science Approvals

<u>Approvals</u>	<u>Dates</u>

Contact Information:

For any questions, please contact:

Vassilis J. Tsotras

Department of Computer Science and Engineering  
Bourns College of Engineering  
University of California Riverside, CA 92521 USA

Phone: 951-827-2888

Email: [tsotras@cs.ucr.edu](mailto:tsotras@cs.ucr.edu)

## **Table of Contents**

<b>EXECUTIVE SUMMARY</b>	<b>5</b>
<b>SECTION 1: INTRODUCTION</b>	<b>6</b>
<b>SECTION 2: PROGRAM</b>	<b>13</b>
<b>SECTION 3: PROJECTED NEEDS</b>	<b>17</b>
<b>SECTION 4: PROGRAM FACULTY AND STAFF</b>	<b>19</b>
<b>SECTION 5: COURSES</b>	<b>20</b>
<b>SECTION 6: RESOURCE REQUIREMENTS</b>	<b>25</b>
<b>SECTION 7: GRADUATE STUDENT SUPPORT</b>	<b>26</b>
<b>SECTION 8: GOVERNANCE</b>	<b>27</b>
<b>SECTION 9: SENATE REGULATION CHANGES</b>	<b>28</b>
<b>APPENDIX A: PROGRAM BYLAWS</b>	<b>30</b>
<b>APPENDIX B: NEW COURSE SYLLABI</b>	<b>34</b>
<b>APPENDIX C: LETTERS OF SUPPORT</b>	<b>44</b>

## **EXECUTIVE SUMMARY**

This document is a proposal for a Master of Science (M.S.) degree in Data Science (DS), which will be jointly managed by the departments of Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE). Degree requirements and administration of the program are described in this document.

Data has become ubiquitous in everyday life, impacting every profession, including manufacturing, logistics, health care, public safety, and the military. Data also permeates all aspects of science, engineering, and other academic disciplines. As a result, the field of Data Science has emerged as a new academic discipline: the study of data itself. Data Science deals with obtaining insight and information from the analysis of large collections of data. The proposed MS in DS is a comprehensive program studying how data can be collected, transformed, analyzed, and used to solve problems across many application areas.

At UCR, relevant courses related to data management, data mining, information retrieval, big data, machine learning, and artificial intelligence have been offered in the Computer Science & Engineering and the Electrical & Computer Engineering Departments. These courses are regularly offered and are very popular. However, our current MS curricula in BCOE do not permit students to obtain a focused mastery of Data Science.

The proposed program will allow students with an undergraduate degree from a quantitative field, some experience in algorithms and software engineering, and an exposure to introductory statistics to enroll in a masters-level program in Data Science that will grant them a broad understanding of the subject, while focusing on parts of it for a deeper understanding depending upon their interests.

The new program will rely on existing faculty and will be built mostly on existing courses (only three new courses will be added) within the two departments. It will leverage upon existing facilities in the two departments. Future course offerings will also be through CSE and ECE and the program faculty will be from these departments.

## **SECTION 1: INTRODUCTION**

### **1.1 Program Objectives**

The objective of the MS in Data Science program is to provide training in various aspects of the data lifecycle. Students will gain exposure to data collection, data cleaning, data integration, data management, and data visualization, as well as the theories and techniques necessary for data analysis from data mining, machine learning, information retrieval, and artificial intelligence.

The program aims to admit students from various backgrounds with undergraduate training in quantitative fields (e.g., engineering, physics, math, statistics). We expect that applicants will have some experience in programming, software engineering, and algorithms, and some exposure in probability/statistics. The committee overseeing the formation of the program has considered this aspect very carefully and designed a program that provides both breadth and depth. Two new courses were designed with this purpose in mind: They introduce students from different backgrounds to the basic tools and theory in the Data Science field. Students will then be exposed to the breadth of the area through a set of core courses. They will also be able to focus on various aspects of data science and gain in-depth knowledge through specific electives. At the end, students will complete a capstone project (new course) where they will combine technical, analytic, and interpretive skills to design and execute a large-scale data science project that has a focus on real-world applications.

It is also possible to accept students whose undergraduate education did not include the expected experience in programming, software engineering etc. Examples are students whose undergraduate degrees are in chemistry, biology, economics or sociology. Such students may still be admitted to the program with the stipulation that they complete missing courses at the undergraduate level at UCR. The CSE and ECE departments are working on a “bridge” program that could be used as a first step by students who need instruction in undergraduate fundamentals, such as programming, algorithms, and data structures, prior to entering graduate programs in CSE, ECE, or Data Science. Through the bridge program, students without the appropriate background can still finish their MS degree in Data Science within 2 years. We expect the bridge program to increase the reach of this Data Science MS program in the near future.



## 1.2 Historical Development of Data Science and Departmental Strengths

We live in a world where data is being generated continuously by scientific experiments, digital processes, sensors, social media, mobile devices, etc. The term “big data” refers to data that is arriving from multiple sources at an alarming volume, velocity, and variety. Data Science is a new field that deals with the management of and extraction of knowledge from big data. As a scientific field, Data Science affects research in many domains, including biological sciences, physical sciences, social sciences, and humanities. The importance of Data Science is evident by various related UC-wide initiatives. As an example, UCB has recently created a separate Data Science Division (<https://data.berkeley.edu/>).

The White House “*Big Data Research and Development Initiative*” committed \$200 million to “extract knowledge and insights from large and complex collections of digital data, accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning.” NIH launched the *Big Data To Knowledge (BD2K)* initiative “to enable biomedical research as a digital research enterprise, to facilitate discovery and support new knowledge.” *Harnessing the Data Revolution* is part of [NSF’s 10 Big Ideas](#). In particular, “Engaging NSF’s research community in the pursuit of fundamental research in data science and engineering, the development of a cohesive, federated, national-scale approach to research data infrastructure, and the development of a 21st-century data-capable workforce.” Other funding agencies (DARPA, IARPA, etc.) have similar research initiatives.

In addition to research, Data Science heavily influences economics and business. Data has become ubiquitous in everyday life: It impacts every profession, from entry-level office workers to CEOs, from team coaches to general managers, from accountants to CFOs. Businesses now have data available to them at a scale that is historically unprecedented; harnessing this data for insight on what customers want provides them with a competitive advantage. Traditional companies (Ford, Walmart, General Electric, etc.) today pride themselves as being transformed to big-data businesses.

Fueled by the explosion of data, Data Science jobs have proliferated and the demand for data scientists is extremely high; moreover, this demand is expected to be strong for years to come. A 2016 McKinsey report forecasted a shortfall of roughly 250,000 data scientists by 2024. Data scientists are the no. 1 most promising job in America for 2019,

## Proposal for M.S. Degree in Data Science

according to a report from LinkedIn. Similarly, according to Glassdoor, a recruiting site, Data Scientist has been the best job in the US (2015-2019) with around 113K median base salary. Three-fifths of the data science and analytics jobs are in the finance and insurance, professional services, and information technology sectors, but the manufacturing, health care, and retail sectors also are hiring significant numbers of data scientists. We thus expect that the new program will be in high demand among students and will serve the UCR community well.

As another indication of the interest in Data Science, we have experienced high demand among graduate students for related courses (Data Mining, Machine Learning, AI, Big Data, etc.) For example, around 70-100 students attended “CS235: Data Mining”, in its last few offerings; similarly “CS236: Database Management”, “CS 226: Big Data Management” and “CS229: Machine Learning” have enrollments around 50-60 students. We believe many of these students would prefer a degree more concentrated on these particular topics, particularly one with a coordinated project to provide hands-on experience. Thus, we feel this proposed MS program will better serve many of our current students.

Preparing the workforce in Data Science is also important for the local community. Here in the Inland Empire, for example, the Naval Surface Warfare Center (NSWC) in Corona has launched the Universal Hub for Big Data, a project to collect and share Navy data, which will require a qualified workforce. Our ability to keep high-tech employers like this in the region depends on our ability to supply professionals capable of satisfying their technical needs. NSWC has recently contacted BCOE expressing strong interest in the proposed MS in DS program.

Finally, the MS in Data Science will be of interest as a career next path, to the UCR students graduating from the newly approved BS in Data Science program. We expect that some of these students will continue to pursue a PhD degree in CSE or ECE. Further, a BS+MS will be a possibility in the future.

We thus believe that the MS in Data Science program will be instrumental in educating the future Data Scientists by building their expertise from solid core knowledge, covering the essentials in managing and analyzing data, as well as covering the applications of Data Science in real life problems.

The CSE and ECE Departments have many faculty that perform research related to Data Science. The existing strength was instrumental in creating recently the Data Science Center. Moreover, through a Data Science Cluster, three more faculty members were hired (Papalexakis, Eldawy, Oymak). Section 4 discusses the initial program faculty (currently 10 CSE and 5 ECE members). There are existing strong research groups working on Big Data, Database Management, Data Mining, Artificial Intelligence, Deep Learning, Time Series, and Vision. Related research is published in the top conferences and journals, and is consistently funded by various grants from NSF, Army, Navy, DARPA and other funding agencies. Graduates from these groups are very much sought after from the industry (including Google, Amazon, LinkedIn, Microsoft, Facebook etc.)

### **1.3 Enrollment Projections**

We believe that the new MS in Data Science will help to increase the overall graduate enrollment in BCOE, which is also a college aim. In Fall 2019, the three MS programs offered by CSE, ECE or both had the following enrollment: CSE 168, ECE 37 and CEN 43 MS students. As of 7/31/20, the number of Fall 2020 SIRs were: CSE 85, ECE 50, CEN 46. We believe that the new MS in Data Science will be at least as popular as the ECE and CEN MS programs.

We thus aim to start with 20 students in the first year of the program and reach a steady state of 50-60 students within 5 years. This would be achieved without hurting enrollment in the CSE, CEN and ECE MS programs (or the new MS in Robotics), since the MS in Data Science offers a different career path than these other MS programs.

Further, we expect that many of these students will stay on for PhDs in CSE or ECE, thus allowing us to select PhD students who have already been at UCR.

### **1.4 Relation to Other Programs in UCR and the UC System**

We note that the MS in Data Science will be a state-supported program focused on students that are interested in the on-campus experience. It is thus different from the self-supported BCOE MSOL program that offers a Data Science specialization (among others).

UCR has recently approved an M.S. in Business Analytics from the School of Business (in collaboration with the Statistics department). This degree is different from our

proposed program as it focuses on non-technical aspects of data management and analysis while we are looking at the computational side of data. UCR's MS in Business Analytics is more equivalent to UCI's MS in Business Analytics.

Within the UC system, UC San Diego has a (self-supported) Masters of Advanced Studies program in Data Science and Engineering that runs over Fridays/Saturdays; this program is offered through their Engineering school. Similarly, UCLA offers through the Engineering School, an on-line Master of Science in Engineering With Certificate of Specialization in Data Science Engineering. UC Berkeley has an on-line M.S. in Information and Data Science that is offered through their School of Information. They further provide the "5th Year Master of Information and Data Science" program, open to Berkeley undergraduate students as a path to earning a professional master's degree in one additional calendar year. UC Berkeley also has a (self-supported) MS in Engineering program through their Electrical Engineering and Computer Science Department, that offers a concentration in Data Science and Systems. UC Irvine offers a M.S. in Business Analytics offered by the School of Business and has a more business rather than a technical focus.

MS in Data Science is offered by many top universities that have strong research in the area. Examples include the Master in Computational Data Science at CMU, the Masters in Data Science at NYU and the Masters in Data Science at Columbia University.

### **1.5 Contributions to Diversity**

Because of its ubiquitousness and inherent interdisciplinarity, Data Science has an enormous, and still largely untapped, potential for increasing diversity in computing. In Fall 2016, women accounted for only 13.8% of the Computer Science and Electrical Engineering undergraduate enrollment (including all majors offered by the departments). Because our program draws on undergraduates from a more diverse set of majors, we expect to have a more balanced set of applicants. Further, due to Data Science's relationship to a large variety of application areas, we expect this major to appeal to a broader set of students (including more women) than a traditional Computer Science or Electrical Engineering degree. The recent surge of workshops and conferences that promote diversity in Data Science and related fields, with prominent examples including "WiML" (Women in Machine Learning; <https://wimlworkshop.org/>), "WiDS" (Women in Data Science; <https://www.widsconference.org/>), and "BPDM" (Broadening Participation in Data Mining; <https://www.facebook.com/BPDMProgram>) is strong empirical evidence for the validity of our premise. Similarly there are initiatives like CAWIT (Center for Advancing Women in Technology; <https://www.cawit.org/>)

whose aim is to increase the participation of women in computing and information technology, by developing new interdisciplinary computing degree programs that educate more women innovators for the Digital Age. CAWIT has recently supported our undergraduate Data Science major with a grant to enable us start and advertise the program.

UCR is an accredited Hispanic Serving Institution (OPEID 00131600), with approximately 35% Hispanic enrollment. This provides an ideal environment for recruitment of underrepresented graduate students. We expect most of the MS in DS students to come from STEM undergraduate programs. Having an MS in DS program will allow us to attract a large number of these students by providing a focus area, thus enriching the diversity of our graduate student pool. The fast-growing nature of the Data Science field is a great motivating factor for these students to complete an MS before entering the workforce. We also expect that some of the MS students will stay on for a PhD, thus enhancing diversity in the associated PhD programs too. The oversight committee of the MS in DS program will organize an open feedback session at the end of each academic year in order to obtain qualitative feedback from students and instructors. In addition, the committee will perform quantitative diversity assessment through anonymous student survey and evaluation, in collaboration with the two participating departments.

### **1.6 Comments from other UC programs**

We identified four Data Science related MS programs in other UCs (all self-supported). We sent copies of this proposal to the chairs of these programs, with a cover letter using the sample provided by the Senate instructions. These programs were (including the date of the letter): (1) UCB: Master of Information and Data Science (9/4/2020), (2) UCB: Master of Engineering in EECS with concentration on Data Science and Systems (9/4/2020), (3) UCSD: Master of Advanced Studies in Data Science and Engineering (9/1/2020), and (4) UCLA: Master of Science in Engineering With Certificate of Specialization in Data Science Engineering (9/1/2020). However, no comments on the proposed program have been received, with three of the programs not responding and the other one stating that this was out of their domain.

### **1.7 Administration of the Program**

The program will be led by a Program Director, assisted by an Associate Director. While the Director will focus on the overall program and coordination among the departments, the Associate Director will serve the role of Graduate Advisor taking care of all graduate student advising issues within the program. A staff member will help the faculty Directors in administering the program. The program faculty will consist of Senate faculty in related research areas from the two departments (see list of the initial program faculty in Section IV). In the interest of efficient administration, a core group of faculty will be appointed to oversee the program and coordinate efforts with the two departments. This Oversight Committee will consist of 5 faculty from the two departments (three from CSE and two from ECE), including the Director and Associate Director.

This initial proposal was created by the following group of faculty:

Samet Oymak (ECE)

Vagelis Papalexakis (CSE)

Mariam Salloum (CSE)

Christian Shelton (CSE)

Vassilis Tsotras (CSE) - Committee Chair

### **1.8 Evaluation of the Program**

As is the norm for all graduate programs at the UCR campus, the program will follow the Senate-mandated review (once every six or seven years). Beginning with the second year, the Program Committee will initiate an internal review of the M.S. Data Science Program.

## **SECTION 2: PROGRAM**

Below we describe the undergraduate admission requirements, the program of study and provide a sample time plan.

### **2.1 Admission Requirements**

All applicants to this program must have completed a Bachelor's degree or its approved equivalent from an accredited institution and to have attained undergraduate record that satisfies the standards established by the Graduate Division and University Graduate Council. Students need experience in a quantitative field with experience in programming, software engineering, algorithms, and background in statistics. Competence in these areas is defined by the following UCR undergraduate courses (or equivalents):

- CS 141 - Intermediate Data Structures and Algorithms
- CS 100 - Software Construction
- MATH 010A - Multivariable Calculus
- MATH 031 - Linear Algebra
- A course covering foundations of probability and statistics (such as STAT 155 - Probability and Statistics for Science and Engineering, or, EE 114 - Probability, Random Variables, and Random Processes in Electrical Engineering)

Applicants who fail to meet this criterion may sometimes be admitted with course deficiencies, provided they take remedial steps to cover the deficiencies. A student who is deficient in a competency area may be asked to complete the corresponding UCR course with a letter grade of at least B, or to pass a challenge examination based on that course's final exam with a grade of at least B. All such remedial work cannot be counted towards the MS degree requirements and should be completed within the first year of graduate study, and in all cases the deficiency(s) must be corrected BEFORE a student can enroll in any graduate course from the same specialty area. The details will be decided by the Graduate Advisor of the program in consultation with the student. The CSE and ECE departments are working currently on a 'bridge' program that can be used as a first step by students who lack basic undergraduate background in programming, algorithms and data structures.

All applicants must submit scores from the Graduate Record Exam, General Test (GRE). Relevant GRE subject tests may be beneficial to the candidate's application, but are not required. Applicants whose first language is not English are required to submit acceptable scores from the TEST of English as a Foreign Language (TOEFL) or the International English Language Testing System (IELTS) unless they have a degree from an institution where English is the exclusive language of instruction. Additionally each applicant must submit letters of recommendation, as per the admission requirements. All other application requirements are specified in the graduate application.

## **2.2 Data Science MS Program**

The MS Data Science program will be built using existing courses, with three new courses<sup>1</sup> (CS/EE 251A: Data Analytics and Exploration, CS/EE 251B: Fundamentals of Data Science and CS/EE 279: Capstone Project in Data Science). The MS in Data Science requires the completion of 49 units of coursework, including a capstone project. There are no thesis or comprehensive exam options; i.e. it falls in the category of a Master's II (with capstone).

Units are divided among core courses (6 courses, for a total of 24 units), elective courses (5 courses, for a total of 20 units), a professional development course (1 unit) and the capstone course. All students must complete the same core courses. Elective courses are selected by the student within a list of possible courses, and students can petition to select a course not on the list.

### **Core courses:**

1. CS/EE 251A: Data Analytics and Exploration (new course)
2. CS/EE 251B: Fundamentals of Data Science (new course)
3. CS 224: Fundamentals of Machine Learning
4. CS 226: Big Data Management
5. CS 235: Data Mining Techniques
6. CS 236: Database Management

### **Elective courses:**

The five electives can be selected from the following two lists of elective courses; at least three of the courses must be from list A. The description of all the elective courses is available later in the proposal. Students may petition for other elective courses; such

---

<sup>1</sup> All three courses have been approved by both the CSE (10/7/20) and ECE (10/14/20) Departments and are in the process of senate approval.



## Proposal for M.S. Degree in Data Science

petitions require approval of the program graduate advisor.

### Elective List A:

1. CS 205: Artificial Intelligence
2. CS 225: Spatial Computing
3. CS 227: Probabilistic Models for Artificial Intelligence
4. CS/EE 228: Introduction to Deep Learning
5. CS 229: Machine Learning
6. CS 242: Information Retrieval and Web Search
7. CS/EE 248: Optimization for Machine Learning
8. EE 231: Convex Optimization in Engineering Applications
9. EE 240: Pattern Recognition
10. EE 244: Computational Learning

### Elective List B:

1. CS 210. Scientific Computing
2. CS 211. High Performance Computing
3. CS/EE 217: GPU Architecture and Parallel Programming
4. CS 234: Computational Methods for Biomolecular Data
5. EE 241: Advanced Digital Image Processing
6. EE 243: Computer Vision
7. EE 250: Information Theory

**Capstone Experience:** Students must complete a capstone course CS/EE 279: Capstone Project in Data Science (new course), under the guidance of the capstone instructor member. The description of the capstone course appears in Section 5.

**Professional Development Requirement:** Students will satisfy the professional development requirement by enrolling in one of the following courses: one quarter of CS 287 (Colloquium in Computer Science), or GDIV 403 (Research and Scholarship Ethics), or at least one unit of CS 298I (Individual Internship).

## 2.3 Other Requirements

There are no field or qualifying examinations. There is no thesis/dissertation or final examination. There are no special requirements over and above the Graduate Division minimum requirements.

## 2.4 Sample Program

Below we provide a sample program. Assuming that a student has no deficiencies and

## Proposal for M.S. Degree in Data Science

is full-time, the normative time from matriculation to degree is 4 quarters. Using the currently planned bridge program in CSE, it is expected that students with deficiencies can still graduate within 2 years.

	Fall	Winter	Spring
Year 1	CS/EE 251A CS/EE 251B CS 236	CS 226 CS 224 Elective	CS235 Elective Elective
Year 2	Elective Elective CS/EE 279		

## **SECTION 3: PROJECTED NEEDS**

### **3.1 Student Demand and Opportunities**

We expect a large demand for the new program. The numbers of students in related programs, like the BS and MS programs in CSE, ECE and CEN continue to increase. The proportion of domestic students in related MS programs is around 17% in CSE, 20% in CEN and 33% in ECE; we expect that for the MS in DS the percentage of domestics will be closer to the ECE example. This is because Data Science seems to be popular with domestic students. Moreover, there are a number of students enrolled in the new B.S. program in Data Science and this program will draw from students who complete the DS B.S. program (the BS in DS inaugural class of Fall 2020 is 100% domestic students). Many students in the various Undergraduate Professional Societies have expressed interest in a Data Science graduate program. While we have most of the courses, the structures of the existing programs do not allow them to take the proper set of courses required for specialized training in Data Science This demand is directly related to opportunities for students after graduation.

### **3.2 Opportunities for Placement of Graduates**

Based on our experience from graduate students working in relevant areas (Databases, Data Mining, Artificial Intelligence, Machine Learning etc.) there is currently very high demand from industry. Moreover, as mentioned in the Introduction, according to Glassdoor, a recruiting site, Data Scientist has been the best job in the US (2015-2019).

### **3.3 Importance to the Discipline**

As a scientific field, Data Science affects research in many domains, including biological sciences, physical sciences, social sciences, and humanities. In addition to the newly approved Data Science undergraduate program, the proposed MS in Data Science allows students to concentrate further on this important field.

### **3.4 Meeting the needs of Society**

Data is an important societal asset. By training more students in Data Science we also create more “citizen scientists”. According to CitizenScience.gov (the official government website dedicated to Citizen Science), a citizen scientist “...participates voluntarily in the scientific process, addressing real-world problems in ways that may include formulating research questions, conducting scientific experiments, collecting

and analyzing data, interpreting results, making new discoveries, developing technologies and applications, and solving complex problems". Such involvement can engage the American public in addressing societal needs and accelerating science, technology, and innovation.

### **3.5 Relation to Research and Faculty Interests**

A critical mass of our faculty are engaged in research and teaching across the full range of areas relevant to the proposed MS program. This is also evident from the fact that almost all the courses for the program already exist at UCR. These areas are already of high interest to faculty. Moreover, faculty is well funded in these research areas.

### **3.6 Program Differentiation**

A current list of Data Science related programs in California appears at <http://datascience.community/colleges>. The majority of these programs are on-line, or focus in Business Analytics. Some are MS programs that provide concentrations in Data Science. Below we describe how the proposed MS in Data Science differs.

The MS in Data Science is a state-supported program. It is thus different from UCR's MSOL program in Data Science as well other similar online programs in other UC campuses. It is also different from MS in Business Analytics or MS in Information Management programs as such programs focus on non-technical aspects of data management and analysis while we are looking at the computational side of data. UCR's MS in Data Science is also different from UC Berkeley's (self-supported) MS in Engineering with Concentration in Data Science and Systems. To get that concentration, students need to take 4 technical courses from a list of approved EECS courses and a capstone project. The UCR MS in Data Science is not a concentration but the whole focus of the MS degree through a well designed curriculum that offers many opportunities to students to train in Data Science related coursework. Our program is also different from the Masters of Advanced Studies program in Data Science and Engineering from UC San Diego (also self-supported), which runs Fridays/Saturdays and focuses on mid-career professionals. It is also different from UCLA's Master of Science in Engineering With Certificate of Specialization in Data Science Engineering, which is similar to our MSOL program.

The MS in Data Science at UCR will be the first state-supported MS program in this important subject.

Among private California Institutions there are two related programs: (1) the MS in Information Systems & Technology with concentration in Data Science, offered by Claremont Graduate University, and (2) the Master of Science in Computer Science (Data Science) offered by USC. We believe that we offer a very competitive program from a public institution that is actually named MS in Data Science.

## **SECTION 4: PROGRAM FACULTY AND STAFF**

The list of the Program Faculty (with a link to their publications) appears below:

### **CSE**

Ahmed Eldawy (Assistant Professor; PhD; <https://dblp.uni-trier.de/pers/hd/e/Eldawy:Ahmed> )

Vagelis Hristidis (Professor; PhD; <https://dblp.uni-trier.de/pers/hd/h/Hristidis:Vagelis> )

Eamonn Keogh (Professor; PhD; [https://dblp.uni-trier.de/pers/hd/k/Keogh:Eamonn\\_J=](https://dblp.uni-trier.de/pers/hd/k/Keogh:Eamonn_J=) )

Paea LePendu (Assistant Teaching Professor; PhD; <https://dblp.uni-trier.de/pers/hd/l/LePendu:Paea> )

Amr Magdy (Assistant Professor; PhD; [https://dblp.uni-trier.de/pers/m/Magdy\\_0001:Amr.html](https://dblp.uni-trier.de/pers/m/Magdy_0001:Amr.html))

Evangelos Papalexakis (Assistant Professor; PhD;

[https://dblp.uni-trier.de/pers/hd/p/Papalexakis:Evangelos\\_E=](https://dblp.uni-trier.de/pers/hd/p/Papalexakis:Evangelos_E=))

C.V. Ravishankar (Professor; PhD; [https://dblp.uni-trier.de/pers/hd/r/Ravishankar:Chinya\\_V=](https://dblp.uni-trier.de/pers/hd/r/Ravishankar:Chinya_V=) )

Mariam Salloum (Assistant Teaching Professor; PhD; <https://dblp.uni-trier.de/pers/hd/s/Salloum:Mariam> )

Christian Shelton (Professor; PhD; [https://dblp.uni-trier.de/pers/hd/s/Shelton:Christian\\_R=](https://dblp.uni-trier.de/pers/hd/s/Shelton:Christian_R=) )

Vassilis Tsotras (Professor; PhD; <https://dblp.uni-trier.de/search?q=tsotras> )

### **ECE**

Salman Asif (Assistant Professor; PhD; [https://dblp.uni-trier.de/pers/hd/a/Asif:Muhhammad\\_Salman](https://dblp.uni-trier.de/pers/hd/a/Asif:Muhhammad_Salman) )

Bir Bhanu (Professor; PhD; <https://dblp.uni-trier.de/pers/hd/b/Bhanu:Bir> )

Samet Oymak (Assistant Professor; PhD; <https://dblp.uni-trier.de/pers/hd/o/Oymak:Samet> )

Amit Roy-Chowdhury (Professor; PhD; [https://dblp.uni-trier.de/pers/hd/r/Roy=Chowdhury:Amit\\_K=](https://dblp.uni-trier.de/pers/hd/r/Roy=Chowdhury:Amit_K=) )

Nanpeng Yu (Assistant Professor; PhD; <https://dblp.uni-trier.de/pers/hd/y/Yu:Nanpeng> )

### **STAFF**

One FTE for administrative support, primarily for graduate student admissions, enrollment and advising. Initial support may be less than 1 FTE, ramping up as the program matures.

### **TEACHING RESOURCES**

The new program is based on existing courses from CSE and ECE. The three new courses will be cross-listed between the two departments which will share responsibilities in teaching them. In the Appendix we include letters of support from the two department chairs that also discuss the sharing of the teaching.

## SECTION 5: COURSES

### Core Courses

#### New Courses Developed for the Program

**CS/EE 251A.** *Data Analytics and Exploration* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): CS141, CS100, Stat 155 or EE114 or equivalent. This course covers important algorithms relevant to the lifetime of data from data collection and cleaning to integration, data mining and analytics. Topics include: sketch algorithms for computing statistics on data streams; mining social graphs, including community detection and graph partitioning; Data Science lifecycle and techniques on data cleaning, data integration, Exploratory Data Analysis, and visualization.

**CS/EE 251B.** *Fundamentals of Data Science* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): Math 010A, Math 031 or EE 020, CS100, Stat 155 or EE114 or equivalent. Explores theoretical tools in data science and their applications in data science. The course introduces and motivates statistical and computational viewpoints on data analysis. Topics include the manipulation of data as vectors, drawing inferences from data as distributions, and quantifying data uncertainty for data analysis. The course will also include in-class and homework exercises on practical applications of these theoretical data science tools.

**CS/EE 279.** *Capstone Project in Data Science* (4) Lecture, 1 hour; outside research, extra readings, 9 hours. Prerequisite(s): Enrollment in Master in Data Science. Co-requisites: CS/EE 251A, CS/EE 251B, CS224, CS226, CS235, CS236. Covers combining technical, analytic, and interpretive skills to design and execute a large-scale data science capstone project that has a focus on real-world applications. Provides an opportunity to integrate all of the core skills and concepts learned throughout the program and prepares students for long-term professional success in the field. Emphasizes collaboration and communication in both written and oral form.

### Existing Core Courses

**CS 224:** *Fundamentals of Machine Learning.* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): CS 100, STAT 155, MATH 31 . A study of generative and discriminative approaches to machine learning. Topics include probabilistic model fitting, gradient-based loss optimization, regularization, hyper-parameters, and generalization. Includes experience with data science programming environments, data from practice, and performance metrics.

**CS 226.** *Big-Data Management* (4) Lecture, 3 hours; term paper, 3 hours.

Prerequisite(s): CS 166. Introduction to the architecture and design of big data management systems. Covers the design of distributed file systems and high throughput databases. Description of popular programming paradigms for big data including MapReduce and Resilient Distributed Datasets. Includes a course project with hands-on experience on open-source big data systems. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 235.** *Data Mining Techniques* (4) Lecture, 3 hours; term paper, 1.5 hours; project, 1.5 hours per week. Prerequisite(s): CS 141, CS 166; CS 170 is recommended. CS 235 online section; enrollment in the online Master of Science in Engineering program. Provides students with a broad background in the design and use of data mining algorithms and tools. Includes clustering, classification, association rules mining, time series clustering, and Web mining. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 236.** *Database Management Systems* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): CS 141; CS 153 or equivalent; CS 166; or consent of instructor. Covers principles of file systems; architecture of database management systems; data models; and relational databases. Also examines logical and physical design of databases; hardware and software implementation of database systems; and distributed databases (e.g., query processing, concurrences, recovery). May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

### **Electives**

**CS 205.** *Artificial Intelligence* (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): CS 170 or equivalent. Examines knowledge representation and automated reasoning and their use in capturing common sense and expert knowledge. Also addresses predicate and nonmonotonic logics; resolution and term rewriting; reasoning under uncertainty; theorem provers; planning systems; and belief networks. Includes special topics in natural language processing, perception, logic programming, expert systems, and deductive databases. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 210.** *Scientific Computing* (4) Lecture, 4 hours. Prerequisite(s): CS 012, MATH 010A; MATH 031 or equivalent; or consent of instructor. Utilizes scientific computing in a specific computer science research area. Provides a foundation for pursuit of further studies of special topics in scientific computing. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 211.** *High Performance Computing* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): CS 161 or consent of instructor. Introduces performance optimization for sequential computer programs. Covers high performance computing on multicore shared memory computers and on distributed memory computing clusters. Also covers



high performance scientific libraries and computing application development using pthreads, OpenMP, and Message Passing Interface (MPI) parallel file systems. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS/EE 217. *GPU Architecture and Parallel Programming*** (4) Lecture, 3 hours; consultation, 1 hour. Prerequisite(s): CS 160 with a grade of “C-” or better or consent of instructor. Introduces the popular CUDA based parallel programming environments based on Nvidia GPUs. Covers the basic CUDA memory/threading models. Also covers the common data-parallel programming patterns needed to develop a high-performance parallel computing applications. Examines computational thinking; a broader range of parallel execution models; and parallel programming principles. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 225. *Spatial Computing*** (4) Lecture, 3 hours; individualized study, 3 hours. Prerequisite(s): graduate standing; or consent of instructor. Introduction to the spatial computing technologies and techniques. Covers the fundamentals, the present, and the emerging use cases of spatial data analysis systems. Topics include spatial data modelling, spatial relationships, storage, indexing, query processing, and recent trends in the field. Includes a research-oriented project and hands-on experience on spatial technologies. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 227. *Probabilistic Models for Artificial Intelligence*** (4) Lecture, 3 hours; written work, 3 hours. Prerequisite(s): CS 141, STAT 155. Covers methods for representing and reasoning about probability distributions in complex domains. Focuses on graphical models and their extensions such as Bayesian networks, Markov networks, hidden Markov models, and dynamic Bayesian networks. Topics include algorithms for probabilistic inference, learning models from data, and decision making. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS/EE 228: *Intro to Deep Learning***. (4) Lecture, 3 hours; written work, 3 hours. Prerequisite(s): CS 225 or EE 236 or EE 231 or EE 244 or CS 171 or EE 142 or consent of the instructor. Explores fundamentals of deep neural networks and their applications in various machine learning tasks. Includes the fundamentals of perception, approximation, neural network architectures, loss functions, and generalization. Addresses optimization methods including backpropagation, automatic differentiation, and regularization. Covers non-standard problems including autoencoders, weak supervision and probabilistic models. Presents applications in machine learning/computer vision.

**CS 229. *Machine Learning*** (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): CS 100, STAT 155. CS 229 online section; enrollment in the Online Master-in-Science in Engineering program. A study of supervised machine learning that emphasizes discriminative methods. Covers the areas of regression and classification. Topics include linear methods, instance-based learning, neural networks, kernel

machines, and additive models. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 234.** *Computational Methods For Biomolecular Data* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): CS 111; CS 141 or CS 218; STAT 155 or STAT 160A. A study of computational and statistical methods aimed at automatically analyzing, clustering, and classifying biomolecular data. Includes combinatorial algorithms for pattern discovery; hidden Markov models for sequence analysis; analysis of expression data; and prediction of the three-dimensional structure of RNA and proteins. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS 242.** *Information Retrieval and Web Search* (4) Lecture, 3 hours; term paper, 1.5 hours; project, 1.5 hours per week. Prerequisite(s): CS 141, CS 166. CS 242 online section; enrollment in the online Master of Science in Engineering program. Introduces Information Retrieval (IR) principles and techniques for indexing and searching document collections with special emphasis on Web search. Includes text processing, ranking algorithms, search in social networks, search evaluation, and search engines scalability. May be taken Satisfactory (S) or No Credit (NC) with consent of instructor and graduate advisor.

**CS/EE 248:** *Optimization for Machine Learning.* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): CS 229 or EE 231 or EE 244 or consent of the instructor. Explores efficient optimization algorithms for machine learning. Emphasizes fundamental principles, provable guarantees, and contemporary results. Includes fundamentals of optimization (first-order methods, stochastic algorithms, accelerated schemes, non-convex optimization, regularization, and black-box optimization). Also covers connections to statistical learning (empirical risk minimization, finite-sample guarantees, and high-dimensional problems).

**EE 231.** *Convex Optimization in Engineering Applications* (4) Lecture, 3 hours; term paper, 3 hours. Prerequisite(s): EE 230. Covers recognizing and solving convex optimization problems that arise in engineering applications. Explores convex sets, functions, and optimization problems. Includes basics of convex analysis, least-squares, linear and quadratic programs, semidefinite programming, minimax, and other problems. Addresses optimality conditions, duality theory, theorems of alternative and applications, interior-point methods, and applications in engineering.

**EE 240.** *Pattern Recognition* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): EE 141 or consent of instructor. EE 240 online section; enrollment in the Online Master-in-Science in Engineering program. Covers basics of pattern recognition techniques. Topics include hypothesis testing, parametric classifiers, parameter estimation, nonparametric density estimation, nonparametric classifiers, feature selection, discriminant analysis, and clustering.

**EE 241.** *Advanced Digital Image Processing* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): EE 152 or consent of instructor. Covers advanced topics in digital image processing. Examines image sampling and quantization, image transforms, stochastic image models, image filtering and restoration, and image data compression

**EE 243.** *Advanced Computer Vision* (4) Lecture, 3 hours; outside research, 3 hours. Prerequisite(s): EE 146 or consent of instructor. EE 243 online section; enrollment in the Online Master-in Science in Engineering program. A study of three-dimensional computer vision. Topics include projective geometry, modeling and calibrating cameras, representing geometric primitives and their uncertainty, stereo vision, motion analysis and tracking, interpolating and approximating three-dimensional data, and recognition of two-dimensional and three-dimensional objects.

**EE 244.** *Computational Learning* (4) Lecture, 3 hours; research, 3 hours. Prerequisite(s): graduate standing or consent of instructor. Explores fundamental computational learning techniques. Topics include elements of learning systems, inductive learning, analytic learning, case-based learning, genetic learning, connectionist learning, reinforcement learning and integrated learning techniques, and comparison of learning paradigms and applications.

**EE 250.** *Information Theory* (4) Lecture, 3 hours; extra reading, 3 hours. Prerequisite(s): EE 215. An overview of fundamental limitations imposed on communication systems. Topics include Shannon's information measures, weak and strong typicality, lossless data compression, source and channel models and Shannon's coding theorems, channel capacity and the rate-distortion function, Gaussian sources and channels, and limits of communication between multiple terminals.

## **SECTION 6: RESOURCE REQUIREMENTS**

All the technical resources required by the M.S. Data Science program are already available in and for the two participating departments including computing facilities, library resources, teaching laboratories and research facilities. The only additional resources would be office space and one FTE for administrative support (initial support may be less than 1 FTE, ramping up as the program matures).

## **SECTION 7: GRADUATE STUDENT SUPPORT**

MS Data Science students are expected to be self-supported. However, GSR and Teaching Assistantships may be available on a case-by-case basis.

## **SECTION 8: GOVERNANCE**

The Program Faculty will consist of Senate faculty in related research areas to Data Science, drawn from the two departments. Program Faculty members shall support the program through instruction of courses, supervision of students, activity in Data Science research, or program administration. All Program Faculty are eligible to vote on matters related to the MS in Data Science Program. All changes to the MS in Data Science Program or curriculum must be approved by a majority of the Program Faculty.

The program will be led by a Program Director, assisted by an Associate Director. The Director is appointed by the Dean of BCOE with consultation from the Program Faculty. The Program Director will rotate among the 2 departments. While the Director will focus on the overall program and coordination among the departments, the Associate Director will serve the role of Graduate Advisor taking care of all graduate student advising issues within the program. A staff will help the faculty Directors in administering the program. The staff will report to the Director and the Director will report to the Dean of BCoE.

A core group of the program faculty (including the Director and Associate Director) will be appointed to form the Oversight Committee, whose task is to oversee the program and coordinate efforts with the departments. The committee will consist of three faculty from CSE and two faculty from ECE.

## **SECTION 9: SENATE REGULATION CHANGES**

No changes in Senate Regulations at the Divisional level or in the Assembly of the Academic Senate will be required.

## **APPENDIX A: PROGRAM BYLAWS**



## **MS in Data Science Program Bylaws**

Creation Date: June 2, 2020

Approval Date:

### **I. Objective**

- A. The MS in Data Science is housed in the Bourns College of Engineering (BCOE), and is a joint program between the departments of Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE).
- B. The objective of the MS in Data Science is to provide training in various aspects of Data Science. Students graduating from the program will gain exposure to the foundational principles underlying the full data lifecycle, from storage to management to analysis.

### **II. Membership**

- A. The faculty associated with the program, called the Program Faculty, is drawn from UCR Senate faculty in related research areas from the CSE and ECE departments.
- B. Program Faculty members shall support the program through instruction of courses, supervision of students, activity in Data Science research, or program administration.
- C. All Program Faculty are eligible to vote on matters related to the MS in Data Science Program.
- D. All changes to the MS in Data Science Program or curriculum must be approved by a majority of the Program Faculty.
- E. UCR Senate faculty outside of CSE and ECE whose research or teaching activities align with the mission of the MS in Data Science are eligible to be Cooperating Faculty in the program. Cooperating Faculty do not have a vote in the program, but are eligible to participate in meetings of the Program Faculty.
- F. Membership Changes
  - 1. Nominations of prospective members to the Program Faculty or Cooperating Faculty may be made by any faculty member in CSE or ECE.
  - 2. New Program Faculty or Cooperating Faculty shall be appointed by a majority vote of the Program Faculty, based on a review of the nomination and the recommendation of the Oversight Committee, defined in III.A below.

3. Members of the Program Faculty may terminate their association with the MS in Data Science Program after so informing the Program Director in writing.
4. Participation as Program Faculty or Cooperating Faculty shall be reviewed every three years to ensure that all members are meeting their obligations to the MS in Data Science Program.

### III. Administration

A. A core group of faculty, called the MS in Data Science Program Oversight Committee, shall oversee the program and coordinate efforts with the departments.

#### B. Composition

1. The Program Oversight Committee is chaired by the Director, or by the Associate Director in the Director's absence.
2. The Program Oversight Committee consists of five (5) members (including the Director and Associate Director), all of whom are members of the Program Faculty.
3. Three (3) faculty from CSE and two (2) faculty from ECE departments shall be on the Oversight Committee. Faculty with joint appointments in multiple departments shall specify the one department they represent.

#### C. Duties

1. The duties of the Director include
  - a. providing overall academic and administrative leadership for the program,
  - b. overseeing the development and implementation of program policies,
  - c. representing the interests of the program to the College, the Campus and University administrators,
  - d. calling and chairing meetings of the program,
  - e. managing the program's budgets,
  - f. ensuring the accuracy of publications related to the program including web pages and catalog copy, and
  - g. coordinating the program's teaching needs with the teaching assignments of the constituent departments.
2. The duties of the Associate director include
  - a. serving as the Graduate Advisor for the MS in Data Science program,
  - b. coordinating administration with the Office of Graduate Studies,
  - c. submitting course change or approval forms, and

d. assisting the Director as needed.

D. Appointments

1. The Dean of BCOE appoints the Director with consultation from the Program Faculty, in a manner consistent with the appointment of other program directors and department chairs. The Director reports to the BCOE Dean.
2. It is expected that Directors should alternate between the two departments. Any exception will require a majority vote of the Oversight Committee.
3. Director appointments are for three (3) years, except when circumstances require otherwise.
4. Members of the Oversight Committee, other than the Director, are nominated and elected by the Program Faculty, in accordance with the provisions of bylaw III.B above.
5. The Associate Director will be appointed by the Director from the membership of the Oversight Committee.

IV. Meetings

A. The Program Faculty

1. The Program Faculty will meet as necessary, but at least once a year.
2. Three or more faculty from the Program Faculty can call a meeting.

B. The Program Oversight Committee

1. The Program Oversight Committee will meet at least once per academic term, on a schedule set by the Director.
2. Three or more faculty from the Program Oversight Committee can call a meeting.

C. Members will be notified of meetings at least a week in advance.

D. A quorum for meetings of the Program Faculty consist of 50% of the Program Faculty.

E. A quorum for meetings of the Program Oversight Committee consist of 4 members of the Program Oversight Committee.

## **APPENDIX B: NEW COURSE SYLLABI**

## CS/EE 251A : Data Analytics and Exploration Spring 2021

**Instructor:** Mariam Salloum / Vagelis Papalexakis

**Contact Info:** [msalloum@cs.ucr.edu](mailto:msalloum@cs.ucr.edu) / [epapalex@cs.ucr.edu](mailto:epapalex@cs.ucr.edu)

### **Credits / Type**

4.0 Units

Lecture: 3 hours

Research (outside): 3 hours

### **Description:**

This course covers important algorithms relevant to the lifetime of data from data collection and cleaning to integration, data mining and analytics. Topics include: sketch algorithms for computing statistics on data streams; mining social graphs, including community detection and graph partitioning; Data Science lifecycle and techniques on data cleaning, data integration, Exploratory Data Analysis, and visualization.

**Prerequisite(s):** CS141, CS100, Stat 155 or EE114 or equivalent.

### **Relevant Textbooks**

- (abbreviated MMD) [Mining of Massive Datasets](#) by Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman
- (abbreviated EDA) [Experimental Design and Analysis](#) by Howard J. Seltman. 2018
- Selected papers (See assigned readings in the schedule)

### **Grading:**

- Homework (x5) 35% (assignments include both a written and programming component)
- Midterm (x2) 40%
- Final Project 25%

### **Tentative Schedule**

Week	Lecture Topics	Readings (Book/Papers)
1	Probability Review. Distributions (single and multi-dimensional), Central-limit Theorem, Expectation, Mean, Variance, Moments,	<a href="https://www.stat.cmu.edu/~hseltman/309/Book/chapter3.pdf">https://www.stat.cmu.edu/~hseltman/309/Book/chapter3.pdf</a> (EDA book Ch.3)
2	Mining Data Streams: sampling, filtering (e.g. bloom filters), sketch algorithms	<a href="http://infolab.stanford.edu/~ullman/mds/ch4.pdf">http://infolab.stanford.edu/~ullman/mds/ch4.pdf</a> (MMDS book Ch. 4)

## Proposal for M.S. Degree in Data Science

3	Sketch Algorithms Cont. (Count-Min and Heavy Hitters)	Notes from <a href="http://theory.stanford.edu/~tim/s15/l12.pdf">http://theory.stanford.edu/~tim/s15/l12.pdf</a>
4	Mining Social Graphs: Intro to Social Networks as Graphs, Clustering (distance measures, Betweenness and Girvan-Newman alg.)	<a href="http://infolab.stanford.edu/~ullman/mds/ch10n.pdf">http://infolab.stanford.edu/~ullman/mds/ch10n.pdf</a> (MMDS book Ch. 10)
5	Mining Social Graphs Cont.: Community detection and Graph Partitioning (Finding Clique, Bipartite Graphs, Partitioning)	NA
6	Data Science lifecycle & Exploratory Data Analysis & Ethics of Big Data	<a href="http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf">http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf</a> (EDA book Ch. 4-7)  Voosen, P., Big-Data Scientists Face Ethical Challenges After Facebook Study. The Chronicle of Higher Education. Retrieved from <a href="https://www.chronicle.com/article/Big-Data-Scientists-Face/150871">https://www.chronicle.com/article/Big-Data-Scientists-Face/150871</a>
7	Data Visualization (including topics such as dimensionality reduction, tSNE)	(t-SNE) <a href="https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf">https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf</a>  (UMAP) <a href="https://arxiv.org/pdf/1802.03426.pdf">https://arxiv.org/pdf/1802.03426.pdf</a>
8	Data cleaning techniques	<a href="http://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf">http://dc-pubs.dbs.uni-leipzig.de/files/Rahm2000DataCleaningProblemsand.pdf</a>
9	Data Integration methods & ETL (schema matching, record-linkage, data fusion)	<a href="#">Big Data Integration</a> by Xin Luna Dong and Divesh Srivastava
10	Data Integration methods cont.	NA

## **CS/EE 251B: Fundamentals of Data Science**

Spring 2021

**Instructor:** Samet Oymak, Christian Shelton

**Contact Info:** [oymak@ece.ucr.edu](mailto:oymak@ece.ucr.edu), [cshelton@cs.ucr.edu](mailto:cshelton@cs.ucr.edu)

### **Credits and type**

4.0 Units

Lecture: 3 hours

Research (outside): 3 hours

### **Course Information**

#### **A. Course Description**

Explores theoretical tools in data science and their applications in data data science. Introduces and motivates statistical and computational viewpoints on data analysis. Topics include the manipulation of data as vectors, drawing inferences from data as distributions, and quantifying data uncertainty for data analysis. Also includes in-class and homework exercises on practical applications of these theoretical data science tools.

**B. Prerequisite(s)** Math 010A, Math 031 or EE020, CS100, Stat 155 or EE114 or equivalent, or permission by instructor

### **Syllabus**

#### **Week 1**

Data as a vector I: motivation for linear algebra in data science, norms of vectors and matrices, eigenvalues and eigenvectors, fundamental subspaces

#### **Week 2**

Data as a vector II: Hermitian and positive semidefinite matrices, singular values, QR decomposition, principal component analysis (PCA), low-rank approximation

#### **Week 3**

Data analysis with linear algebra: least-squares, pseudo-inverse, condition number, ridge regression, in-class exercise on MNIST dataset and PCA

#### **Week 4**

Data as a distribution I: motivation for statistics and probability in data science, the randomness in data, random variable, conditional probability, expectation, variance, moments

#### **Week 5**

## Proposal for M.S. Degree in Data Science

Data as a distribution II: covariance matrices, correlation coefficient, data normalization, multivariate Gaussians, law of large numbers, in-class exercise on analyzing covariance matrices on the Adult dataset

### **Week 6**

Inference with data: Parameter estimation, unbiased estimator, bias-variance decomposition, maximum likelihood estimator (MLE), maximum a posteriori estimation (MAP), log likelihood

### **Week 7**

Applications of Estimation: in-class exercises on MLE in clinical data, minimum mean-square error (MMSE), prediction with least-squares, coefficient of determination, in-class exercise on MMSE in time series prediction

### **Week 8**

Quantifying uncertainty with data: hypothesis testing, confidence intervals, p-value, Student's t-test, bootstrapping, in-class exercise on hypothesis testing on the movie ratings

### **Week 9**

Optimization with data: the role of data in modern optimization problems, loss functions, convexity, gradient, in-class exercise on gradient descent and least-squares on the Adult dataset

### **Week 10**

Overflow: Finish the material from earlier weeks or practice for the final exam.

### **Textbooks and Related Materials**

Recommended sources:

1. Textbook for linear algebra topics: Gilbert Strang, "Linear Algebra and Learning from Data", Wellesley-Cambridge Press, 2019.
2. Textbook for statistical topics: Larry Wasserman, "All of Statistics: A Concise Course in Statistical Inference", Springer, 2013. [available as a free pdf from <https://link.springer.com/content/pdf/10.1007%2F978-0-387-21736-9.pdf>]

### **Grading** TBD

Participation 5%

HWs 40% (mix of coding projects and problem solving on paper)

Midterm 25%

Final 30%



## **CS/EE 279 : Capstone Project in Data Science**

Fall 2022

**Instructor:** Mariam Salloum

**Contact Information:** [msalloum@cs.ucr.edu](mailto:msalloum@cs.ucr.edu)

### **Credits/Type**

4.0 Units

Lecture: 3 hours

Research (outside): 3 hour

### **Short Description (<= 50 words)**

Covers combining technical, analytic, and interpretive skills to design and execute a large-scale data science capstone project that has a focus on real-world applications. Provides an opportunity to integrate all of the core skills and concepts learned throughout the program and prepares students for long-term professional success in the field. Emphasizes collaboration and communication in both written and oral form.

**Prerequisites:** Enrollment in Master in Data Science.

**Co-requisites:** CS/EE 251A, CS/EE 251B, CS224, CS226, CS235, CS236.

### **Course Objectives**

At the end of this course, students will be able to demonstrate their knowledge, skills and abilities to develop and execute a data science project using real-world data and effectively communicate their results to a technical and non-technical audience.

Students will be able to:

- Formulate a research question, problem or hypothesis that can be answered or tested using real-world data;
- Collect and manage data to devise solutions to their research question, problem or hypothesis;
- Select, apply and evaluate models, tools and methods to address their research question, problem or hypothesis. This includes building an end-to-end analysis pipeline covering data sourcing, cleaning/preparation, integration and transformation, and visualization;
- Interpret and assess their results and evaluate the limitations of their findings;
- Prepare a professional report of their work and effectively communicate their findings to a technical and non-technical audience.

### **Grading**

Students will work on a quarter-long project in teams of 2-3 students. The grading rubric is focused on group and individual project representations, project report, and a final web-based deliverable. In addition to these assignments, students are evaluated based on their participation in class discussions, and by their group-mates based on contributions to the group.

- 5% - Class participation (class discussions) and weekly meetings with course instructor
- 50% - Project Deliverables
  - Proposal (due Week 2) - Project proposal
  - Phase 1 (due Week 4) - Code and status report
  - Phase 2 (due Week 7) - Code and status report
  - Phase 3 (due Finals Week) - Code, and final report
- 35% - In-class presentations - instructor evaluation and peer feedback on presentations
- 10% - Web-based final deliverable

### **Readings**

There is no textbook for this course. Readings are drawn from various relevant books, articles and academic papers that are available online.

### **Schedule**

#### **Week 1 - Introduction**

##### ***Topics***

- Reviewing the data science life-cycle
- Case studies of organizations using “big data” effectively
- Project and group selection

##### ***Required Readings***

- Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins and Nina Kruschwitz, Big data, analytics, and the path from insights to value. MIT Sloan Management Review (2010).  
<https://sloanreview.mit.edu/projects/analytics-the-new-path-to-value/>
- Robert, Christian. 2020. “The 9 Pitfalls of Data Science.” *Chance* 33. [UCR Library]
- (optional) Press. G. (2013). A very short history of data science. Forbes. Retrieved from  
<https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#7cfe337655cf>

## Week 2 - Data Science Applications

### *Topics*

- Data anonymity
- Selected readings from DS applications, focused on areas such as Social Media Analysis, Social and Information Networks, Healthcare and Medicine

### *Required Readings*

- Anna Monreale, Gennady Andrienko, Natalia Andrienko, Fosca Giannotti, Dino Pedreschi, Salvatore Rinzivillo, Stefan Wrobel. Movement Data Anonymity through Generalization, Transactions on Data Privacy, 2010. <http://www.tdp.cat/issues/tdp.a045a10.pdf>
- Selected readings from papers highlighted in <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9912/9874>
- (optional) Voosen, P., Big-Data Scientists Face Ethical Challenges After Facebook Study. The Chronicle of Higher Education. Retrieved from <https://www.chronicle.com/article/Big-Data-Scientists-Face/150871>

## Week 3 - Data Science Ethics

### *Topics*

- Skills for collecting, storing, sharing and analyzing data derived from human subjects including data used in algorithms and examining ethical implications.

### *Required Readings*

- 1 - Data Skeptic
  - O'Neil, Cathy. (2013) On being a data skeptic. p. 1-19. Sebastopol, CA: O'Reilly Media. <http://www.oreilly.com/data/free/files/being-a-data-skeptic.pdf>
  - boyd, danah. (2017)
  - Ethical side effects of the publish or perish system: p-hacking and small sample size Gelman, Andrew and Loken, Eric. (2014) The statistical crisis in science. American Scientist. Accessed online: <https://www.americanscientist.org/issues/id.16259,y.2014,no.6,content.true?page.1,css.pri nt/issue.aspx>
- 2- Data Sharing / Ethics
  - Libby Bishop (2017). Big data and data sharing: Ethical issues. UK Data Service, UK Data Archive. Accessed online: [https://bigdata.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing\\_ethicalissues.pdf](https://bigdata.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethicalissues.pdf)
- 3- Building fair systems/ Ethics
  - Toward accountability: Data, Fairness, Algorithms, Consequences. Data and Society: Points. [blog post] Accessed online: <https://points.datasociety.net/toward-accountability-6096e38878f0>

- Crawford, Kate. (2013, April 1) The hidden biases in big data. Harvard Business Review. <https://hbr.org/2013/04/the-hidden-biases-in-big-data> Lerman, Jonas. (2013, September)
- Big data and its exclusions. Stanford Law Review. Accessed online: <https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusion>

#### **Week 4 - Group Presentations I**

##### ***Topics***

- Group Presentations

##### ***Required Readings***

- NA

#### **Week 5 - Communication and Storytelling**

##### ***Topics***

- Power of storytelling and narrative
- Tactics for presenting and sharing information

##### ***Required Readings***

- Selected readings from [Interactive storytelling : 7th International Conference on Interactive Digital Storytelling, ICIDS 2014, Singapore, Singapore, November 3-6, 2014 : proceedings](#)

#### **Week 6 - Data Visualizations**

##### ***Topics***

- How people and organizations process information and make decisions
- Use of data visualization for communication

##### ***Required Readings***

- Selected readings from [2019 IEEE Visualization in Data Science \(VDS\)](#) IEEE Visualization in Data Science (Conference) (2019 : Vancouver, B.C.)
- Gavett, G. (2014) How data visualization answered of retail's most vexing questions. Harvard Business Review. Retrieved from <https://hbr.org/2014/05/how-data-visualization-answered-one-of-retails-most-vexing-questions>

#### **Week 7 - Group Presentations II**

##### ***Topics***

- Group Presentations

***Required Readings***

- NA

**Week 8 - New trends / topics in Data Science**

***Topics***

- Highlight current research work in data science

***Required Readings***

- Selected readings from KDD, ICML, VLDB, IEEE Big Data

**Week 9 - Guest presentations**

***Topics***

- Guest speaker will discuss their experience in industry

***Required Readings***

- NA

**Week 10 - Final Group Presentations and Deliverables**

***Topics***

- Prepare for final in-class group presentations
- Deliver final presentations and submit project deliverables

***Required Readings***

- None

## **APPENDIX C: LETTERS OF SUPPORT**

# Proposal for M.S. Degree in Data Science



**Computer Science  
and Engineering**

Walid A. Najjar  
Professor and Chair  
Department of Computer Science and Engineering  
351 Winston Chung Hall Riverside, CA 92521  
Tel 951.827.5639 Fax 951.827.4643  
[najjar@cs.ucr.edu](mailto:najjar@cs.ucr.edu) [www.cs.ucr.edu/~najjar](http://www.cs.ucr.edu/~najjar)

August 27, 2020

To Whom It May Concern:

This letter is in strong support for the proposed Master of Science (MS) program in Data Science at UCR, to be jointly offered by the Departments of Computer Science & Engineering and Electrical and Computer Engineering.

Data Science has grown out of the need to integrate computational and statistical approaches to processing and interpreting data. Tools originating from data science are now becoming indispensable in today's science, technology, and business, fueling the demand for data scientists. Recognizing this need, our department has taken the initiative to develop research and educational programs in Data Science at UCR.

In collaboration with other departments on campus, an online MS program in Data Science is already being offered. This fall we are expecting the inaugural class of the new undergraduate program in Data Science (offered in collaboration with the Statistics Department). Recently the Data Science Center has been established that includes multiple newly hired faculty members, and has been given designated space in the new MRB building. Creating a state-supported MS program in Data Science is the next step in this endeavor.

This program will address critical and documented shortage of highly trained college graduates with an advanced degree in Data Science, in industry, government, and academia.

The CSE Department enthusiastically supports the creation of the Data Science MS program and is fully committed to providing necessary resources within its capabilities for the instruction and advising of its students.

A handwritten signature in blue ink that reads "Walid A. Najjar".

Walid A. Najjar

Professor and Chair  
Department of Computer Science and Engineering  
Bourns College of Engineering  
University of California Riverside



# Proposal for M.S. Degree in Data Science



Department of Electrical and Computer Engineering  
343 Winston Chung Hall  
900 University Avenue  
Riverside, CA 92521

August 30, 2020

To  
Academic Senate:

Dear Members of the Academic Senate:

It is my pleasure to provide the strongest possible support for the MS in Data Science program. This program will be housed in the Bourns College of Engineering, and is cross-disciplinary, across the departments of Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE). It will draw upon courses from the existing programs from the departments, including three new cross-listed courses.

Data Science is strategically and technically a very important area that studies how to obtain insight and information from the analysis of large collections of data. As data has become ubiquitous in everyday life, it impacts every profession, including manufacturing, logistics, health care, public safety, and the military. Data is also important in all aspects of science and engineering. The proposed MS in Data Science is a comprehensive program studying how data can be collected, transformed, analyzed, and used to solve problems across many application areas. Students will acquire the cross-disciplinary breadth required for this important and emerging field and can focus, through electives, on specific areas of interest. The proposed program does so at very little expense, since the teaching and research infrastructure are already in place.

ECE expects to interact extensively with the proposed MS in Data Science program by participating in teaching the required and elective courses, in data science research and the mentoring of students through projects and advising, and in helping with the program administration. The program will contribute in a great many positive ways to the ECE department.

In summary, I am extremely supportive of this program and believe it will greatly benefit the students and will help raise UCR's profile. Please do not hesitate to contact me should there be any questions. Sincerely,

Amit Roy-Chowdhury  
Professor and Bourns Family Faculty Fellow  
Chair, [Electrical and Computer Engineering](#)  
University of California, Riverside

Tel 951.827.2484 • Fax 951-827-2425 • [www.ece.ucr.edu](http://www.ece.ucr.edu)

*This letter is an electronic communication from UC Riverside, a campus of the UC system.*



# Proposal for M.S. Degree in Data Science



Office of the Dean  
900 University Avenue  
446 Winston Chung Hall  
Riverside, CA 92521

8/26/2020

To whom it may concern:

I am writing this letter in enthusiastic support for the enclosed proposal to establish a Master of Science degree program in Data Science. This program will be jointly administered within BCOE jointly by the departments of Electrical and Computer Engineering and Computer Science and Engineering. I have had detailed conversations with Professor Tsotras and the program committee and fully support the academic program and administrative structure. I commit to working with them to insure the program's success.

This program will help address the critical and documented shortage of college graduates educated in Data Science and the critical interpretation and analysis of large datasets. We expect students attracted to this program to come from a variety of backgrounds and other interests, increasing the diversity among Engineering students, and those in computational fields in particular.

The Bourns College of Engineering looks forward to launching this MS Data Science degree program. It is an important part of keeping our curriculum current and educating our students.

Sincerely,

A handwritten signature in black ink, appearing to read "Chris Lynch".

Prof. Christopher S. Lynch  
William R. Johnson Jr. Family Chair  
Dean, Bourns College of Engineering  
University of California, Riverside

Tel 951.827.5190 • Fax 951-827-3188 • [www.engr.ucr.edu](http://www.engr.ucr.edu)

*This letter is an electronic communication from UC Riverside, a campus of the UC system.*

Professor Jason Stajich, Chair  
Riverside Division

May 3rd, 2022

Dear Prof. Stajich,

Please find below our response to the comments from the Committee on Diversity, Equity, and Inclusion and the CNAS Executive Committee on our proposal for a MS in Computational Data Science. We would like to thank the committee members for their careful consideration of our proposal.

Sincerely,



V. Tsotras

on behalf of the MS in Computational Data Science Program Committee

## **Response to the Committee on Diversity, Equity, and Inclusion**

In summary, the CODEI committee: (1) asked for ways to ensure that student survey data would remain anonymous/not be compromised by the demographic information collected in surveys, and, (2) suggested partnering with organizations related to underrepresented groups in data science.

We would like to clarify that the demographic data will not be collected by our surveys. Instead we will use data collected by Institutional Research (together with their application, admission and retention data). Our surveys will be optional and will focus on assessing the student's self-efficacy in various topics / technical tools and whether they perceive they have sufficient opportunities as part of the program (opportunities can include mentorship, internships, summer research, etc.) To create and administer the surveys we will use an outside company and such surveys are approved by IRB. As the company holds the survey responses we do not have access to who responded and how. Because the demographic data will be collected in a separate survey from the educational component, there will not be a way of cross-correlating the information.

With respect to suggestion item (2), we note that we have already established collaborations/partnerships with the URM organizations mentioned in the proposal through existing research projects/grants. We will use these partnerships to advertise the program and attract URM students. Examples include local HSIs (like CSUSB, CalPoly-Pomona, California Baptist University, La Sierra University and all local community colleges), the Council for the Advancement of Black Engineers (CABE), the Computing Alliance for Hispanic Serving Institutions (CAHSI) network as well as the local WIDS conference (Women in Data Science).

## Response to the CNAS Executive Committee

The CNAS EC response raised four points:

1. The committee noted the changed title. This title change (Computational Data Science) was a direct response to previous concerns that the program would block other similar efforts. We made the degree title more specific to allow for other programs to concentrate on other aspects of Data Science. We are not against such programs and have specifically changed the title of this proposed program to make it clear that such programs could exist. We have also made it clear to the Statistics Department that we would not stand in the way of any analytics-oriented Data Science Masters (and would help in its creation if requested).
2. CNAS EC stated that the diversity section needed to be more specific. The reviewed proposal had 13 pages of details that were developed specifically for this program. We have addressed CODEI's remaining minor points above.
3. The response states a "worry about the unintended consequences of duplication of learning goals and materials across the colleges." The courses for this program have already been approved. If this worry relates to duplication of program-level goals, its focus on the computational aspects delineates it from other potential programs (as noted in point 1). Other broad disciplines have specialized graduate programs, focusing on specific areas, as this proposal does. Our pedagogical goal is to provide students who have a computational undergraduate degree a Masters program that focuses on the computational aspects of Data Science, including data storage, indexing, and collection, and the algorithmic considerations for its processing. With a 4-quarter program, there is not time for the breadth necessary to cover all aspects of Data Science.
4. The response raised general concerns about revenue sharing. These are not points that can be solved or addressed by this proposal. Nor should such concerns restrict the ability of a proposal to move forward.



## *Academic Senate*

### **EXECUTIVE COUNCIL**

April 28, 2022

**To:** Victor Rodgers  
Chair, BCOE Faculty Executive Committee

Vassilis Tsotras  
Professor, Department of Computer Science and Engineering

**From:** Jason Stajich  
Chair, Riverside Division

**RE:** **2nd Round-Proposal for a Master of Science Degree in Computational Data Science**

Dear Victor and Vassilis,

I write to provide the consultative feedback regarding the subject proposal that I trust will prove helpful for a potential revision.

I highlight for your attention the comments attached from the Committee on Diversity, Equity, and Inclusion and the CNAS Faculty Executive Committee.

Sincerely,  
*/s/Jason*

**COMMITTEE ON DIVERSITY, EQUITY, & INCLUSION**

April 15, 2022

To: Jason Stajich  
Riverside Division Academic Senate

From: Katherine Stavropoulos, Chair  
Committee on Diversity, Equity, & Inclusion

Re: [Campus Review] 2nd Round-Proposal for a Master of Science Degree in  
Computational Data Science

The DEI committee reviewed the 2nd Round-Proposal for a Master of Science Degree in Computational Data Science document, which addresses issues of diversity previously raised by the committee. However, it remains unclear how the proposers will undertake a quantitative diversity assessment through anonymous student surveys. Depending on the projected size of the program, asking students to provide information about their gender/race/ethnicity may undermine efforts to anonymize the survey and could lead to either: a) students not feeling comfortable disclosing concerns, or b) a lack of participation from students. It would be helpful to have more information on how the proposers plan to complete these surveys (e.g. what questions do they plan to ask, how will they promote transparency/participation from students, etc).

Furthermore, the list of organizations meant to increase the participation of underrepresented groups in data science is great. It would be even better if the proposed MS program made it a goal to contribute to these programs or advertise within them. The grant mentioned might already be a part of that, but it isn't very clear from what is written.



7 April 2022

To: Jason Stajich, Chair  
Riverside Division

From: Theodore Garland, Jr., Chair, Executive Committee  
College of Natural and Agricultural Science

Re: Campus Review New Master Degree Proposal: 2nd Round-Proposal  
for a Master of Science Degree in Computational Data Science

The CNAS Executive Committee does NOT support this proposal: not a single member voted in favor.

We were unimpressed by the revisions.

Changing the title was the main thing noted. Other edits were non-substantive.

The section on diversity seems boilerplate and needs to be more detailed/specific.

We worry about the unintended consequences of duplication of learning goals and materials across the colleges.

We understand that the current budget model does not encourage cross-college collaborations, especially when a revenue-generating program is being developed. This is a structural problem with the budget model, and needs to be brought to the attention of the Provost and Chancellor.

Cheers,

A handwritten signature in black ink that reads "Ted Garland".



## *Academic Senate*

### **GRADUATE COUNCIL**

April 21, 2022

To: Jason Stajich, Chair  
Riverside Division

From: Don Collins, Chair  
Graduate Council

Re: [Campus Review] New Master Degree Proposal: 2nd Round-Proposal for a  
Master of Science Degree in Computational Data Science

Graduate Council reviewed the revised proposal for a Master of Science Degree in Computational Data Science at their April 21, 2022 meeting and approved this version of the proposal.



## *Academic Senate*

### **PLANNING & BUDGET**

April 12, 2022

To: Jason Stajich, Chair  
Riverside Division

From: Dana Simmons, Chair  
Committee on Planning and Budget

**RE: [Campus Review] New Master's Degree Proposal: 2nd Round-Proposal for a Master of Science Degree in Computational Data Science**

Planning & Budget (P&B) reviewed the revised proposal for a new Master of Science degree in Computational Data Science at their April 12, 2022 meeting. P&B supported this revised version of the proposal that now includes a letter of support from BCOE Dean Lynch.