PROPOSAL FOR AN UNDERGRADUATE PROGRAM LEADING TO A

Bachelor of Science (B.S.) Degree in

Data Science and Analytics

Proposal Submitted by

Alex Petersen

Associate Professor

Management of Complex Systems

School of Engineering

Proposal Development Team

Christian Fons-Rosen, Associate Professor, Economics and Business Management

Paul Maglio, Professor, Management of Complex Systems and Cognitive and Information Sciences

Russ McBride, Assistant Professor, Management of Complex Systems

Andrew Shaver, Assistant Professor, Political Science

Jeff Yoshimi, Associate Professor, Cognitive and Information Sciences

DRAFT: November 9, 2021

Proposed Major in Data Science & Analytics

1. Program Description and Rationale

1.1 Overview

There is increasing demand in academia and industry for computationally adept analysts capable of extracting value from data by integrating into teams tasked with large-scale workflows[1] – including collecting data from a myriad of existing databases and other less structured data sources; integrating these data into common data formats; analyzing and modeling multi-variable relationships; and communicating via graphical visualization or by designing interactive web platforms that strategically facilitate data-driven decision making. In addition to technical know-how, analysists should be grounded in data ethics associated with collection, ownership, privacy and systemic bias that are of prevailing concern regarding human-data based research and industrial data platforms.

The proposed undergraduate major in Data Science and Analytics (DSA) aims to train students in this multi-dimensional T-shaped skillset – one that integrates deep computational capabilities with a broad transdisciplinary understanding of societal context, real-world applications, and ethical best practices. Such horizontal integration of soft data-communication skills and technical know-how will complement other educational programs, both on campus and in the region, which are specialized in highly theoretical approaches, such as computer science, or domain-specific approaches, such as business analytics. Equipped with a T-shaped skillset, students will be competitive for a wide range of data-oriented industry positions, as well as competent in the analytic toolset fundamental to graduate education in a wide range of STEM and interdisciplinary programs. To prepare students for a range of complex real-world problems, the proposed DSA educational framework integrates People, Planet and Profit (PPP) sustainability perspectives[2] (see Figure 1) by way of flexible application-domain electives that leverage existing courses and faculty strengths in three broad areas: *Understanding and Modeling People*; *Environment and Sustainability*; *Policy and Decision Making*.

Graduates of this program will address Central Valley workforce development needs – whether by seeding an information-based economy through the development of a local supply of data scientists and related start-ups, or by developing remote sensing and other techno-informatic management capabilities critical to *Precision Ag*.[3] As a reflection of the campus' diverse student body, graduates will provide underappreciated perspective[4] to the nation's increasingly data-intensive workforce.[5]

To address this competitive workforce demand nationally, an increasing number of universities offer undergraduate DSA programs. Two lists of undergraduate data science programs, one complied in June 2019[6] and the most recent US News 2021 ranking,[7] indicate a variety of DSA programs ranging from certificates, to minors, degree concentrations, and standalone bachelor degrees. Given the

[1] Börner, K., et al. (2018). Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. Proceedings of the National Academy of Sciences, 115(50), 12630-12637.
[2] Savitz, A. (2013). The triple bottom line: how today's best-run companies are achieving economic, social and environmental success-and how you can too. John Wiley & Sons.
[3] National Academies of Sciences, Engineering, and Medicine, 2018. Improving Data Collection and Measurement of Complex Farms. National Academies Press.
[4] National Academies of Sciences, Engineering, and Medicine, 2018. Minority serving institutions: America's underutilized resource for strengthening the STEM workforce. National Academies Press.
[5] Information technology and the US Workforce: Where are we and where do we go from here?. National Academies Press, 2017; Promising practices for strengthening the regional STEM workforce development ecosystem. National Academies Press, 2016; Strengthening Data Science Methods for Department of Defense Personnel and Readiness Missions. National Academies Press, 2017; Future US workforce for geospatial intelligence. National Academies Press, 2013.
[6] https://www.collegekickstart.com/blog/item/undergraduate-schools-with-strong-data-science-programs
[7] US News 2021 Best Undergraduate Data Science Programs; Notably, this ranking includes "Data Analytics/Science" as a distinct program category, with only 19 entries.

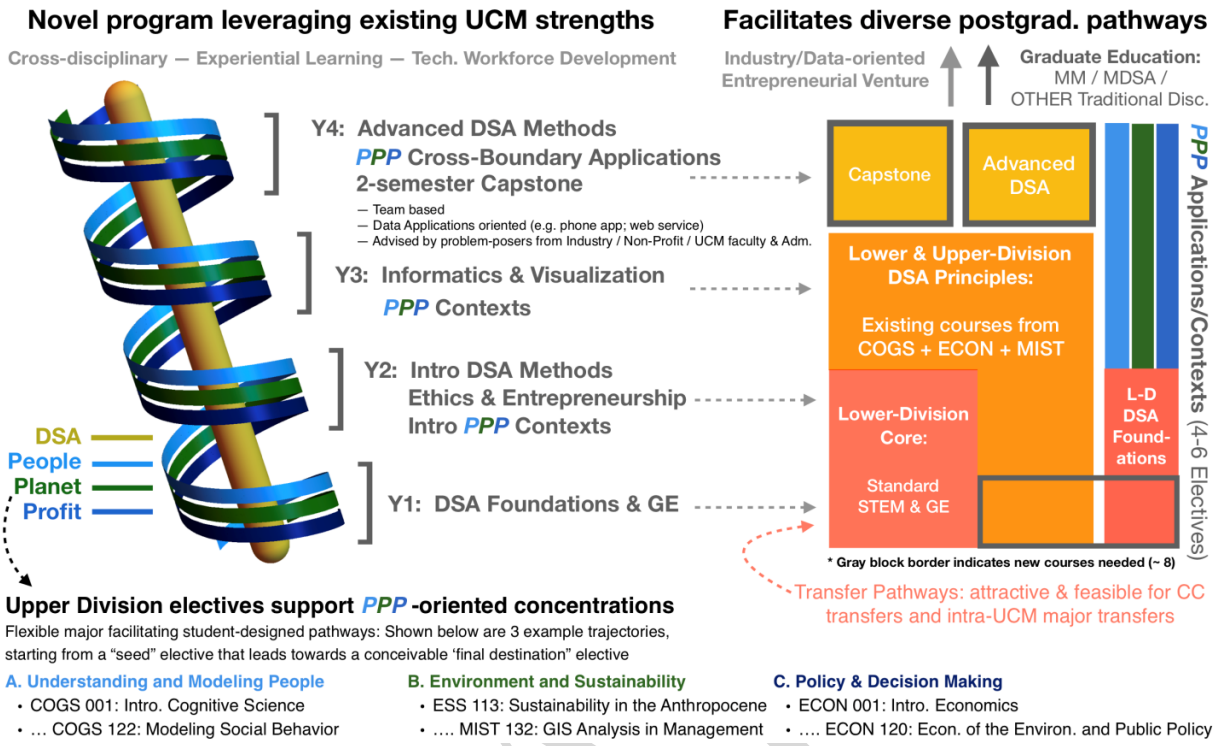


Figure 1. DSA educational framework at the nexus of People, Planet and Profit (PPP) sustainability perspectives, thereby integrating around the campus' mission and existing faculty strengths. (Left) PPP electives provide context for data-driven education, inquiry and real-world applications. Students in Year 1 (Y1) complete foundational methods courses and GE requirements, similar to other undergraduate STEM degree programs – thereby facilitating efficient transfer in and out of the proposed program to other STEM majors; Y2 introduces computational methods around relevant PPP contexts, as well as introductory entrepreneurship to guide translating ideas into practice; Y3 introduces core themes of informatics and visualization, providing the opportunity for students to specialize with select PPP electives; Y4 introduces advanced methods relevant to job market and graduate school demand, and orients students towards cross-boundary DSA applications, in particular via an integrative team-based Capstone project (2-semester) to address a real-world problem defined by an industry advisor or local UCM faculty or even a campus administrative unit (e.g. the Library or Facilities Management). (Right) Vertical Gantt chart shows curriculum comprised mostly of existing courses. Only 5 to 8 new courses would be needed to launch the DSA program (those indicated by a thick block border). (Bottom) Students choose ~ 4-6 upper-division electives that enrich their understanding of the endless frontier of DSA applications.

abundance of DSA-oriented industries and research universities in California, we anticipate robust initial enrollment and sustainable growth. Because this is a relatively new type of undergraduate degree, we are limited in available comparison programs, even within the UC. However, University of Michigan offers a relevant example, having launched *two* "Data Science" programs in 2016, thereby demonstrating the appeal for both deep and broad cross-disciplinary approaches to delivering DSA education at the undergraduate level. Another example of a public university offering multiple data science pathways to appeal to different student backgrounds and needs, is Penn State University, which offers *three* Data Science bachelors programs, one in each of three schools.

In support of the campus' mission to support interdisciplinary educational approaches, the proposed DSA program will leverage a unique transdisciplinary configuration of faculty from four departments – Economics and Business Management (EBM), Cognitive and Information Sciences

(CIS), Management of Complex Systems (MCS), and Political Science (POLISCI) – to support the campus' stated mission to "foster and encourage cross-disciplinary inquiry and discovery". Given the program's focus on integrating People, Planet and Profit (PPP) sustainability perspectives, the convergence of these departments is natural and befitting other transdisciplinary data science programs.[8,9] Faculty from these departments are well-versed in the technical domains of ethics, information science, inferential methods, and systems thinking – and as such, are prepared to educate students in applied data science,[10] including best practices of data-driven visualization and decision-making.[11] The proposal was developed jointly by the four departments in the context of the proposal to establish the new E & J Gallo School of Management. The DSA major is intended to start with the new school in AY 2023 or AY 2024.

A relevant value proposition of the program for the campus is that it is faculty FTE-neutral, relying on existing faculty; moreover, it has relatively small resource requirements owing to the low material cost of delivery (leveraging students' personal laptops as primary computing device and free open-source code and data repositories). Given the proliferation of online certificate programs in data science methods, another important value proposition and comparative advantage of the proposed program is the experiential learning gained through an integrative senior year capstone, where teams will apply DSA methods to solve real-world challenges identified by campus research advisors and industry partners.

1.2 Program contribution to undergraduate education at UCM

The Data Science and Analytics (DSA) Major was developed by four departments working together – Economics and Business Management (EBM), Cognitive and Information Sciences (CIS), Management of Complex Systems (MCS), and Political Science (POLISCI) in consultation with other UC Merced faculty. This program and development team also dovetails with the Masters of Data Science and Analytics (MDSA) that has been proposed and has received positive review by the campus' Graduate Committee and Senate Committees (CAPRA, CoRT, D&E, and LASC) and the Administration (EVC/Provost and the Interim VPDGE). As such, the present proposed DSA major is being designed as part of a robust pipeline for DSA education at UCM, one that leverages multi-disciplinary approaches that are increasingly prevalent in higher education and embraced by the UC Merced research and education mission.

In addition to integrating data-oriented courses from these departments, in particular existing courses comprising the Management Analytics and Decision-Making minor, the program leverages existing lower-division courses from the Applied Mathematics or Computer Science and Engineering undergraduate curricula, which thereby facilitates transfer pathways from Cal State and community colleges, as well as between other STEM majors at UCM. As such, the program starts with traditional STEM approaches to analysis and computing and then integrates real-world application contexts, thereby blending engineering, social and natural science approaches. As such, this interdisciplinary education will equip undergraduate students with the cross-domain breadth, valuable soft (data communication) skills, and technical data-analytics know-how. With this toolkit, graduates will be able to identify and design data-oriented solutions to a wide range of real-world

[8] Why Colleges Are Offering Data Science Programs, US News, 2020.
[9] In particular, MIT offers a novel undergraduate major in *Computer Science, Economics and Data Science*: see Two Sciences tie the knot, MIT News, 2017; Degree description.
[10] National Academies of Sciences, Engineering, and Medicine. (2018). Envisioning the data science discipline: the undergraduate perspective: interim report. National Academies Press; National Academies of Sciences, Engineering, and Medicine. (2018). Data science for undergraduates: Opportunities and options. National Academies Press.
[11] Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. Proceedings of the National Academy of Sciences, 116(6), 1857-1864.

problems, priming them to pursue a diverse panel of post-graduate pathways in academia and industry.

The DSA major will focus on the statistical, computational, and social components of data analysis. The program will combine solid mathematical and statistical foundations with an applied mindset to bring the student as near as possible to real-life data science situations. On top of this, aspects like data ethics, data-services entrepreneurship, and problem-driven data challenges encountered in areas like climate change or social justice will be at the forefront, thereby emphasizing the expertise in these domain areas across campus. Lower and upper-division courses satisfy nearly all of the GE requirements; and because of the program's flexible upper division structure, students will gain contextual knowledge by way of self-selected application-domain electives. Consequently, the program may be an attractive choice for students pursuing a double major.

We are gearing the program such that students will graduate from UCM equipped with a solid understanding of two complementary dimensions - *data science* (oriented around measurement, critical reasoning and modeling) and *data analytics* (oriented around data services, applications and decision-making). Student should be able to come up with data-related questions, know how to analyze them with data to finally provide rigorous and quantitative answers to the problems raised, while also identifying limitations to the data-driven approach. This toolkit should allow the student to pose and address questions in very broad fields of study, including agriculture, environment, law, innovation, management, social science, among others.

The degree will be clearly differentiated from existing majors. Compared to the Economics major, as well as the Computer Science and Engineering (CSE) major and Applied Mathematics major with *Computational and Data Sciences Emphasis*, the proposed DSA curriculum will be less focused on theoretical concepts and methods, and instead more geared towards integrating context and methods in the empirical analysis of natural, engineered, and socio-economic systems. Compared to the Cognitive Science major offered by the CIS department, which focuses on the dimensions of human reasoning and behavior, the proposed DSA program is focused more generally on informatics applications for decision-making.

The program will prepare students for both existing and emergent professions in data science and data analytics in the private sector, international organizations, and governmental agencies across the globe. The skills learned in this major can help students in their goal of becoming data managers, entrepreneurs, evaluating colleague's data analytics work at their job, academics in data-related social and life science fields, school and college educators, or data journalists, among many others.

1.3 Job market demand, graduate education, professional school prospects for majors

Professional Job Market

Despite the emerging digital skills gap, starting as early as K-12 education and well-documented in the present data economy,[12] a 2017 report[13] identified DSA as one of the most rapidly expanding pathways leading to careers in finance, professional services, IT and other sectors. While about 40% of these jobs would require an advanced degree, such as a master's in data science and analytics, the remainder could be filled by graduates with a bachelor's degree in DSA. 98% of hires have a

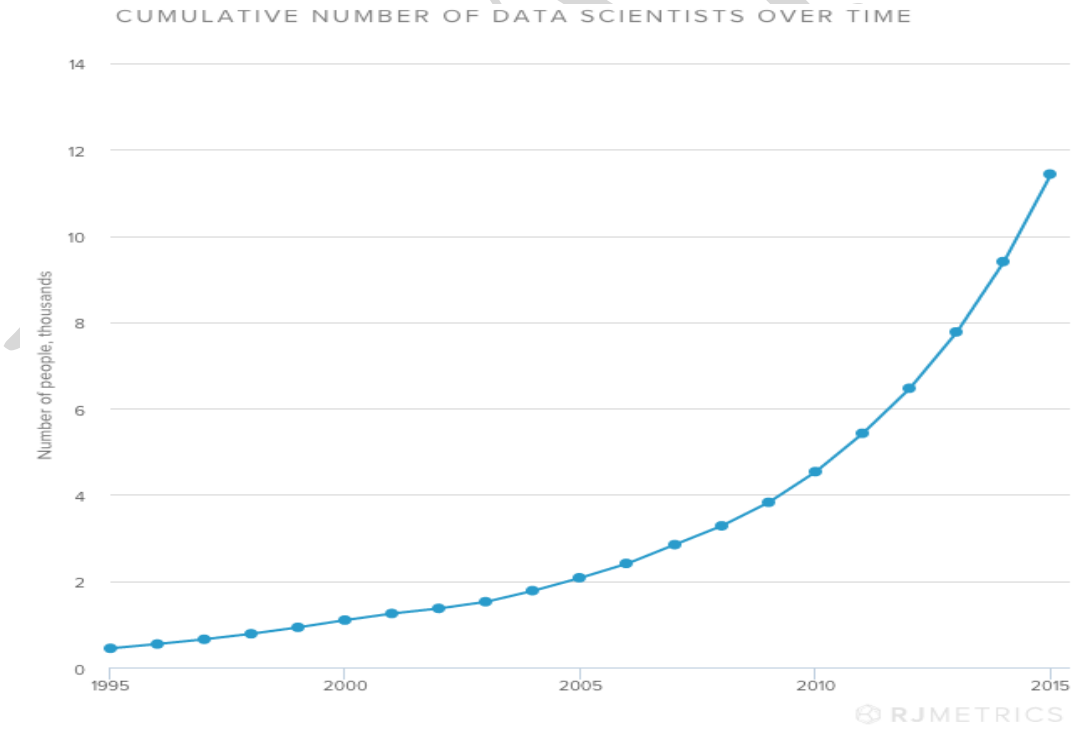
[12] Cappelli, P., 2012. Why good people can't get jobs: The skills gap and what companies can do about it. Wharton Digital Press; Jackman, J. A., et al. (2021). Addressing the digital skills gap for future education. Nature Human Behaviour, 5(5), 542-545; Börner, K., et al. (2018). Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. Proceedings of the National Academy of Sciences, 115(50), 12630-12637.
[13] The Quant Crunch: How the demand for data science skills is disrupting the job market, Burning Glass Technologies, 2017.

minimum of a bachelors degree. In addition to strong historical demand featuring competitive starting salaries, DSA-oriented industries are also anticipated to be resilient to present economic and labor market challenges.[14]

According the LinkedIn's 2020 Emerging Jobs Report, DSA jobs have been growing at roughly 37% annually and are part of the fastest three growing jobs categories (which also includes behavioral health professionals and engineers).[15] Data scientists earn an average of $121,189 annually. Interestingly, three of the five cities with the highest average pay for data scientists are all in California, with the nearby Bay Area leading the salary ranking[16]:

1. San Francisco average DSA salary, $170k, adjusted for cost of living, $128k
2. Boston average DSA salary, $135k, adjusted for cost of living, $120k
3. San Diego average DSA salary, $142k, adjusted for cost of living, $119k
4. Chicago average DSA salary, $116k, adjusted for cost of living, $107k
5. Los Angeles average DSA salary, $129k, adjusted for cost of living, $106k

Figure 2 shows demand for data scientists increasing exponentially over time, as the number of professionals entering the labor market who refer to themselves on LinkedIn as "data scientists" increases with each year.[17] More recently, a LinkedIn report from 2018 identified 151,000 unfilled data science jobs.[18]



<u>Figure 2.</u> Accelerating demand for DSA professionals. Image sourced from whatsthebigdata.com.[17]

[14] Say goodbye to six-figure starting salaries – with these exceptions, CNBC, 2020.
[15] Berger, G. (2020, January). LinkedIn 2020 Emerging Jobs Report. https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/emerging-jobs-report/Emerging_Jobs_Report_U.S._FINAL.pdf Retrieved from URL.
[16] Rayome, A. DeNisco. (Aug 12, 2019). Data scientists earn the highest salaries in these 5 cities. https://www.techrepublic.com/article/data-scientists-earn-the-highest-salary-in-these-5-cities/ Retrieved from URL.
[17] Gilpress (Nov 8, 2015). Top skills and backgrounds of data scientists on LinkedIn. https://whatsthebigdata.com/2015/11/08/top-skills-and-backgrounds-of-data-scientists-on-linkedin/
Retrieved from URL.
[18] LinkedIn Workforce Report | United States | August 2018. Retrieved from URL.

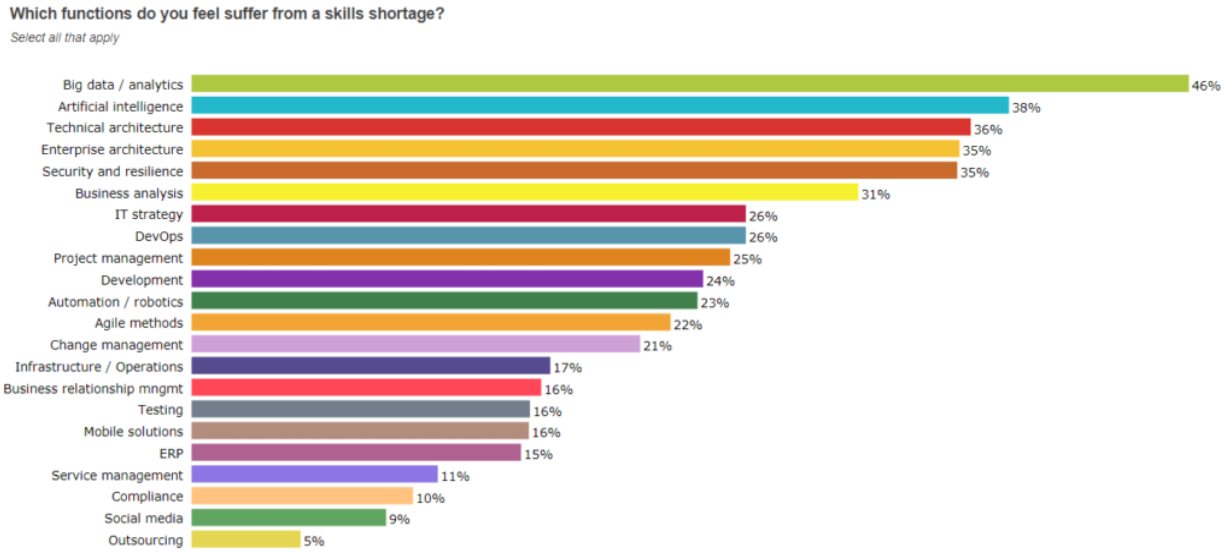


Figure 3. Demand by DSA skillset. Image sourced from datanami.com .[19]

Interestingly, Figure 3 indicates that the greatest demand among the various subfields within the broad field of data science appears to be general DSA analytic skills and the ability to handle big data, the top of the skill list below (as opposed to A.I., robotics, and other sub-fields); another prominent skills vacancy is business analysis, the 6th skill below.[19]

The demand for DSA-trained students appears to be increasing. The same trends are occurring in Europe. According to a report from ITPro, "Simply put, there aren't enough data scientists to go around – the demand for data analysis has grown exponentially over the last few years, and there aren't enough people being trained to meet the demand. This growing gap between demand and available talent has meant that almost half of all European companies are thought to be struggling to fill their data scientist positions". 47% or organizations are struggling to fill their data science positions.[20] According to the Senior Vice President of Global Data Product Management at PepsiCo, Tony Ralph (personal communication), "we simply can't hire data scientists fast enough". (Ralph taught himself data science and worked his way through Silicon Valley until he was earning $1.2M annually running a data science ad analytics team for Walmart).

Hence, the job market demand is booming in the corporate sectors, typically used to improve software, better target advertising efforts, or better analyze customer use, the demands for those with DSA training are extending well beyond typical use cases and into the humanities fields, non-profit organizations, environmental and sustainability work, and policy decision-making, among other areas (c.f., Townsend, 2021, e.g).[21]

Graduate education, professional school prospects

We anticipate that this major would accommodate students pursuing various STEM minors, and possibly even double majors, from across campus. As such, given the interdisciplinary composition

[19] Woodie, A. What's driving data science hiring in 2019? https://www.datanami.com/2019/01/30/whats-driving-data-science-hiring-in-2019/ Retrieved from URL.
[20] ITPro Team. (2019, Sept 17). Data scientist jobs: Where does the big data talent gap lie? https://www.itpro.co.uk/careers/28929/data-scientist-jobs-where-does-the-big-data-talent-gap-lie Retrieved from URL.
[21] Townsend, K. (2021, June 29). Why, how, and what of data science for social impact. https://data.org/news/why-how-and-what-of-data-science-for-social-impact/ Retrieved from URL.

of the curriculum, providing a broad T-shaped in PPP application/contexts via 4-6 electives, complemented by a highly multi-disciplinary composition of the faculty, we believe that graduates of the program would be well-prepared to continue onto interdisciplinary graduate programs, such as the Environmental Systems and Management of Complex Systems PhD programs at UC Merced.

1.4 Expected student demand – Outcomes in similar programs

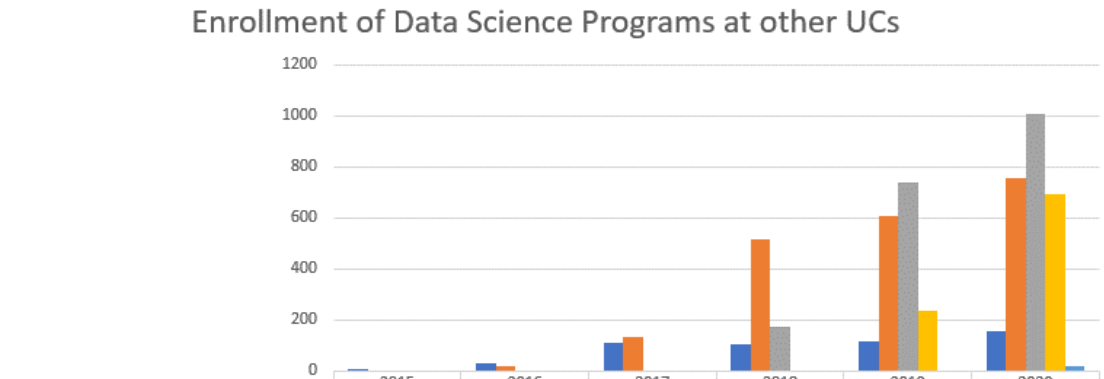
The undergraduate Data Science major is a relatively new paradigm. Hence, there are only a few relevant data sources to draw upon regarding enrollment growth at public universities. This reality identifies the opportunity for UC Merced to establish a prominent program in this domain and leverage an early-mover advantage.

University of California

Five UC campuses currently offer undergraduate degrees in Data Science in AY 2020-21. UC Berkeley offers a Data Science major through the College of Letters & Science where enrollment has exceeded one thousand students in just the first three years from fall 2018 to fall 2020 (data link). Notably, the UCB program is comprised of 28 Domain Emphases (or tracks), that are oriented around three pillars: Computing, Data Science, and Society. By comparison, the proposed UCM DSA program is designed around three tracks that leverage the campus' existing strengths in applied sustainability themes (PPP), and so in this regard, is less focused around a technical computing and algorithm design curriculum common to other computing-oriented programs.

UC Irvine Donald Bren School of Information and Computer Sciences offers a Data Science B. S., which commenced in AY 2015-16; for fall 2020 UCI recorded 157 student-major enrollments (data link). The UC San Diego Halıcıoğlu Data Science Institute launched an undergraduate major in AY 2017, which graduated its first undergraduate cohort of 49 students in AY 2019-20 (data link).

At UC Santa Barbara, the Department of Statistics and Applied Probability oversees both a B.A. and a B.S in Statistics and Data Science with the difference being that the B.A. requires less upper-division units and allows flexibility in courses for the major. The program was first offered fall 2019



	2015	2016	2017	2018	2019	2020
UCI: Data Science, B.S.	4	29	111	106	113	157
UCSD: Data Science, B.S.		16	131	517	609	759
UCB: Data Science, B.A.				171	739	1007
UCSB: Statistics and Data Science, B.A./B.S.					237	696
UCR: Data Science, B.S.						18

Figure 4. Enrollment of Data Science Programs at University of California Locations. Data are obtained from individual Planning Analysis and Institutional Research web page for UCI, UCSD, UCB, UCSB, and UCR .

with an enrollment size of 237 students (data link). UC Riverside is the most recent campus to offer a B.S. in Data Science beginning in the fall 2020 term. The degree program is a collaboration between two colleges, Bourns College of Engineering and College of Natural and Agricultural Sciences, and had an initial enrollment of 18 students. (data link)

University of Michigan

To demonstrate enrollment growth at another public university, Figure 5 shows total academic-year course enrollment numbers from the University of Michigan, which is home to two "Data Science" majors (I.e., of the same name), offered by two different colleges – the College of Literature, Science and the Arts (LSA) and the College of Engineering. These programs thus offer a complementary perspective on enrollment growth, despite having a common framing and name. Since launching in 2016, annual counts of student course enrollments shows that the major offered by the U. Michigan LSA program grew faster than its Engineering counterpart, and also features greater gender balance. The approach Michigan took also illustrates that in a field as broad as data science there is room, and indeed demand, for multiple perspectives and approaches to it.

Anticipated UC Merced Enrollment

Based upon the data from UC campuses, as well as from other public universities, there is demonstrated opportunity for expanding enrollment in the broad domain of DSA. Undergraduate data science programs are no longer found just at the margins of elite private and flagship public universities, but are becoming standard. By way of example, recent data compiled by the *California Alliance for Data Science Education* shows 65 California institutions developing or having already established programs in this direction.

We project that the proposed DSA program could readily attract cohorts of ~50 incoming undergraduate majors within 5 years of commencing. This is based upon the empirical growth of similar UCM STEM programs, according to real 5-year growth rates (calculated over F2015-F2020) for undergraduate declared majors (official UCOP data sourced from here): Computer Science & Engineering (CSE) shows a 73% real growth (i.e. in excess of baseline student population growth); Applied Mathematical Sciences (AMS) grew at 52%; Environmental Systems Science (ESS) grew at

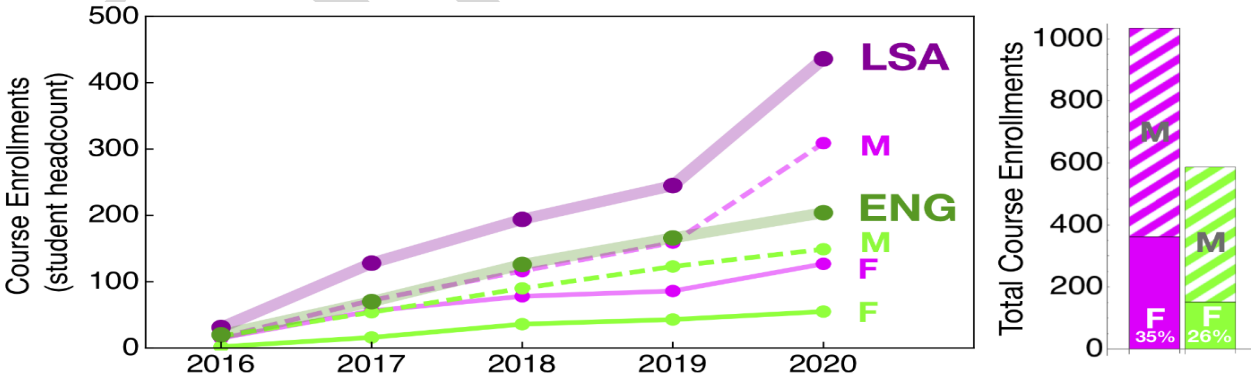


Figure 5. Counts of student-course enrollments shown as Academic Year totals – for two undergraduate "*Data Science*" programs offered by the University of Michigan (official records from the UM Office of the Registrar – "Field of Study" data link). School of Engineering program in Green; School of Literature, Science and the Arts (LSA) program in Magenta. (Left) Enrollment growth over time: Thin solid - Female; Thin dashed - Male; Thick solid - Total (F+M). Counts are total number of students enrolled in data science courses tallied across Fall, Winter, Spring and Summer terms; hence, dividing by 3 approximates the total number of individual students enrolled in the major. (Right) Total enrollment by program and gender group – the LSA program has higher enrollment and higher gender parity.

32%. In terms of student headcounts, over the same period CSE grew from 526 to 1,080 students, corresponding to a 105% increase; AMS grew from 80 to 148 students (85% increase); and ESS grew from 37 to 61 students (65% increase). Hence, there is strong growth in similar STEM programs on campus.

Given the projected growth of campus student enrollment (with the stated Strategic Plan goal of 15,000 students by 2030, representing a roughly ~90% increase over 2019 undergraduate enrollments), and also the steady growth of DSA-related programs on other campuses, it seems reasonable to expect 200-300 total declared majors within 5 years of launch.

1.5 Relation to other majors on campus

The proposed DSA program integrates social science (cognitive, economic, complexity, team science) and design-thinking (engineering, management) perspectives with data-oriented capabilities, with the objective of providing students a holistic training in the sources, applications, and ramifications of data. As such, the proposed DSA program complements two existing undergraduate computer and applied mathematics programs on campus, which are oriented around technical aspects of computer systems and quantitative analysis.

Relation to the Computer Science and Engineering Major

The Computer Science and Engineering (CSE) major is comprised of a rigorous curriculum in core computer programing, algorithms, database, architecture, and operating systems. As an engineering major, students must also complete foundational coursework in engineering physics areas, such as statics and dynamics, circuits, fluid mechanics, and thermodynamics.

As such, there is little overlap in the educational mission of the proposed DSA program and the CSE major, aside from the computer serving as the educational vehicle. In terms of increasing enrollment in fundamental CSE courses, we envision that the proposed DSA program would increase demand for just two lower-division courses – CSE 015 Discrete Math and CSE 019 Intro to Computing. Both these courses are offered in Fall and Spring semesters, and thus could likely accommodate additional enrollment that would not exceed the standard levels of demand associated with other STEM programs on campus. In discussions with the CSE department at their faculty meeting in Spring 2021, they suggested that if enrollment in DSA outpaced their ability to staff CSE 015, the DSA program could teach a dedicated of CSE 015 section. This is a scalable approach to staffing this and other courses, such as CSE 019, that the DSA-affiliated faculty would be open to in order to achieve a win-win configuration.

Despite having different educational missions, we nevertheless anticipate complementarities between the different programs, such as the possibility for cross-listing upper-division courses that are suitable as electives, and increased number of 'hackathon' events to stimulate and integrate the campus' broader computing community.

Relation to the Applied Mathematical Sciences Major - Computational and Data Sciences (CDS) Emphasis Track

The Applied Math CDS emphasis track requires students to select 3 out of 5 specific courses (12 units) that introduce advanced computational methods associated with differential equations, stochastic (random number) modeling, advanced statistics, linear algebra and optimization. These courses are all upper division, and so enrollment impact (in either positive or negative sense) is anticipated to be small, as the upper division MATH courses cover advanced topics that only a small number of select DSA students would likely seek to take.

Moreover, the proposed DSA program is not planning to introduce new courses in any of these areas, but would certainly accommodate students from this and other majors seeking to enroll in DSA courses. As with the CSE program, we anticipate many possible synergies that leverage the computational expertise of the Applied Math faculty, whether it be in cross-listing courses, joint hackathon events, or joint industry speaker days.

The proposed DSA program will ask for an enrollment agreement with Math department that is no different than other STEM programs, drawing upon their lower-division foundations courses (e.g. the Calculus series). Just two courses (MATH 011, 012) will be lower-division requirements of the proposed DSA curriculum. The distinction between the Applied Math CDS track and the proposed DSA Major are thus illustrated by comparing the respective program PLOs. The proposed DSA PLOs are outlined in Section 2.3 (link); for convenience, we provide the Applied Math PLOs below (program description link):

1. Solve mathematical problems using analytical methods.
2. Solve mathematical problems using computational methods.
3. Recognize the relationships between different areas of mathematics and the connections between mathematics and other disciplines.
4. Give clear and organized written and verbal explanations of mathematical ideas to a variety of audiences.
5. Model real-world problems mathematically and analyze those models using their mastery of the core concepts.

2. Program Requirements

2.1 Lower and upper division course requirements

Course Requirements:

Nearly all the required DSA courses already exist. Only one new course would be needed to start the program in AY 2022-23; and only six courses total would need to be introduced for the first cohort to graduate in AY 2026-27. Short descriptions of new courses to-be proposed are provided and indicated with an asterisk (*).

Major – 21 Courses, 80 Units

Lower Division – 8 Courses, 32 Units

Major Preparation Course – 6 Courses, 24 Units

MATH 011*: Calculus I* (4 Units) - Introduction to differential and integral calculus of functions of one variable, including exponential, logarithmic and trigonometric functions, emphasizing conceptual understanding and applying mathematical concepts to real-world problems (approximation, optimization).

MATH 012*: Calculus II* (4 Units) - Continuation of MATH 011. Introduction to integral calculus of functions of one variable and differential equations, emphasizing conceptual understanding and applying mathematical concepts to real-world problem.

***CSE 015**: Discrete Math* (4 Units) - Explores basic concepts of discrete mathematics used in computer science and other disciplines that involve formal reasoning. Topics include logic, proof, counting, discrete probability, relations, graphs, trees, and Boolean algebra.

***CSE 019**: Intro to Computing* (4 Units) - Presents the basics of programming to a student with no prior experience. Concepts of Input/Output, Data Types, Variables and Arrays will be introduced in the context of solving problems. Elementary programming skills such as conditional and loops execution will be emphasized. Object-oriented Programming, recursion and data manipulation will also be introduced. Students are expected to solve problems using different programming paradigms.

ECON 001 or COGS 001

> ***ECON 001**: Introduction to Economics* (4 Units) - Introduction to economics principles and methods, including microeconomics (operation of the economy at the individual and firm level) and macroeconomics (nature and functions of the national economy in a global context).
>
> ***COGS 001**: Introduction to Cognitive Science* (4 Units) - An introduction to the interdisciplinary field of cognitive science. Basic issues related to cognition, including perception, memory, language, learning, problem solving, spatial cognition, attention, mental imagery, consciousness, brain damage, development, and artificial intelligence, are considered from the perspectives of psychology, philosophy, computer science, and neuroscience.

PHIL 002 or PHIL 003

> ***PHIL 002**: Introduction to Ethics* (4 Units) - Consideration of basic questions and themes in moral philosophy through the study of historical and contemporary philosophical texts. Questions may include: Are moral rules relative? What is the best sort of human life? What sorts of acts are right? Should wealth be shared with those less fortunate?
>
> ***PHIL 003:** Contemporary Moral Problems* (4 Units) - Consideration of topics in applied ethics, which may include euthanasia, abortion, economic justice, world hunger, the treatment of animals, and punishment. This course will also cover some more general approaches to ethical thinking, such as theories based on rights and entitlements.

DSA Lower Division Core Course – 3 Courses, 12 Units

***ECON 010**: Statistical Inference* (4 Units) - Introduction to the application of social scientific methods to the study of economics, politics, and management. Covers research design, random sampling, descriptive and inferential statistics, hypothesis testing, and the linear regression model with an emphasis on applications.

***MIST 050**: Introduction to Entrepreneurship* (4 Units) - Provides hands-on training in the practice of entrepreneurship where students learn to make actual money in challenging exercises in "the real world" (or measurably improve a social good if students choose a social entrepreneurship project).

**** DSA 001**: Foundations of Data Science* (4 Units) - Introductory overview course for DSA majors, open to non-majors and counting for the GE Quantitative Reasoning badge, introducing students to principles, contexts and career paths. Transfer students will need to take this course. Pre-requisites: none.

Upper Division - 12 Courses, 44 Units

DSA Principles - 6 Courses, 24 Units

***ECON 110**: Econometrics* (4 Units) – Introduction of problems of observation, estimation and hypotheses testing in economics through the study of the theory and application of linear regression models, critical evaluation of selected examples of empirical research and exercises in applied economics. Pre-Reqs: (ECON 010 or equivalent exam) and (MATH 011 or MATH 021 or equivalent exam).

* ***COGS 1XX/ PHIL 1XX**: Data Ethics (4 Units)* – Course short description in preparation. Pre-reqs: PHIL 002 or PHIL 003 or equivalent.

**** DSA 100**: Foundations of Quantitative Analytics* (4 Units) – A computationally-oriented introduction to essential concepts of statistics, probability, linear algebra and differential equations for DSA majors, and open to non-majors; we will design it to satisfy the UG Quantitative Reasoning badge. Transfer students: Can be satisfied with a similar intro-level Statistics and Probability course. Pre-reqs: Math 012 or equivalent, DSA 001.

***MIST 135:** Technical Communications and Visualization* (4 Units) - Theory of effective communication in quantitative contexts drawing on various presentation modes – written, oral, graphical, powerpoint, poster. Seminars on theory and best-practices combined with individual and team project presentations. Opportunity to improve inquiry-oriented communication skills, to give and receive constructive feedback, and to learn graphical methods for developing striking visualizations.

Advanced Method Courses:

MIST 130 or MIST 134

> ***MIST 130**: Statistical Data Analysis and Optimization in R for Decision Support* (4 Units) – Introduces statistical analysis and optimization for decision support using the R programming environment. Analyze and visualize data and model relationships using graphing techniques, correlation and regression analysis, scenario development and analysis, sensitivity analysis, simulations, and optimization. Emphasizes applied work using real data from resource management and service management examples to support decision making. Case driven and team-focused, emphasizes best practices and professional ethics for analyzing, modeling and communicating empirical data and model results. Pre-Reqs: (MATH 032 or MATH 018 or BIO 018 or equivalent exam) and (MATH 050 or ME 021 or BIOE 021 or CSE 020 or equivalent exam). Pre-reqs to be modified
>
> ***MIST 134**: Methods of Data & Network Science* (4 Units) - Explores methods to efficiently manage and analyze large complex datasets. The computer lab will introduce the Python programming language with a focus on extracting data from websites, exploring, and inquiry-oriented analyses and visualizations. Introduction to network science theory useful in understanding and managing complex socio-technological systems. Pre-Reqs: (MATH 012 or equivalent exam) and (MATH 050 or CSE 020 or ME 021 or BIO 021 or equivalent exam).

DSA 1X1 or DSA 1X2

> *** *DSA-1X1**: Machine Learning & Natural Language Processing* (4 Units) – Course short description in preparation. Pre-Reqs: DSA 001 and ECON 010 or by instructor approval

*** *DSA-1X2:*** *Interactive Data Visualization* (4 Units) – Analyzing and presenting data visually has become one of the most essential skills for students who want to work in data science and related fields. Information Visualization teaches students how to design compelling interactive visualizations of complex data to understand, discover, and present the information. The course combines theoretical knowledge and practical work to develop a well-rounded set of skills to ideate, design, implement, and evaluate sophisticated data visualization projects. The theoretical component aims to provide a mental model to think about the visualization design space in a principled manner. This includes the theory of visual encoding, human perception, and visualization techniques. The practical component aims to teach the skills needed to develop compelling interactive data visualizations for analysis and presentation. Pre-Reqs: MIST 135 or by instructor approval

Application-domain Electives – 4 Courses, 16 Units The DSA curriculum supports gaining broad knowledge of data-oriented application domains by way of a flexible set of electives – requiring a minimum of 4 courses from a list of qualified application-domain courses. Students seeking to include courses not included in this list will be able to submit a written request to be considered for approval by the program's executive committee on a case-by-case basis.

This framework facilitates students specializing in a particular application domain if they plan their electives carefully and early from Y1 onward. In some cases, students will need to identify the pre-requisite courses needed to reach a final destination course, e.g. taking MATH 021 and 022 instead of MATH 011 and 012. While we do not expect any or many students to identify specialized tracks, given the cross-disciplinary composition of the DSA faculty and curriculum, it is worth highlighting the potential for such ambitious schedules that are reminiscent of self-designed majors found at other universities.

Capstone – 2 Courses, 4 Units

*** *DSA 120:*** *Capstone I* (2 Units) – Integrative team-based 2-semester Capstone project to address a real-world problem defined by an academic or industry advisor.

**** DSA 121:*** *Capstone II* (2 Units) – Continuation of DSA 120. Satisfies GE Culminating Experience badge. Pre-Reqs: DSA 120

2.2 Programmatic support for diversity

By designing a project-based and experiential learning-based curriculum, the proposed DSA major will support diversity by naturally drawing on the perspective and prospective of each student, supporting them with an empowering skillset that draws on their individual background, experiences, and passion. Team-oriented projects will be a common course component, which will immerse students in dynamic environments that promote awareness and support for diversity, equity, and inclusion. Awareness and support for DEI will be further supported through data-driven inquiry that exposes the value of diversity in various socio-economic contexts, while also exposing political and algorithmic origins of systemic bias that nevertheless persist in many data-generating socio-technical systems.

By leveraging an organic blend of STEM, social science, and engineering approaches, we anticipate a healthy gender balance. This expectation is supported by enrollment data from the University of Michigan's two data science majors (see Section: Expected Student Demand – Outcomes in Similar

Programs), with the program offered by the multi-disciplinary School of Literature, Science and the Arts (LSA) having greater gender balance than its counterpart offered by the School of Engineering.

2.3 Program Learning Outcomes

Upon completion of the program, students will have the skillset needed to develop a workflow for harnessing and making use of data in a range of real-world contexts and be prepared to engage in lifelong learning and continuing education by leveraging this skillset in either industry or academic future pathways.

Six PLOs for the proposed DSA Major

1. **Methods.** Students will integrate appropriate quantitative, statistical, analytical, algorithmic and coding paradigms to identify knowledge management, planning and strategic decision-making solutions in different organizational and socio-technical situations.

2. **Communication**. Students will selectively draw on different modes of communication (verbal, oral, graphical, code) to inform, engage and inspire in a clear and concise manner to diverse audiences comprised of experts and non-experts.

3. **Design.** Faced with a problem-oriented case, students will design resourceful and ethical informatics-based solutions by integrating data, methods, and web-based technologies to strategically organize, manage, communicate, and deliver information-based services.

4. **Team.** Faced with a problem-oriented capstone project, students will gain experience integrating knowledge, skills, theory, and methods by leveraging team- and data-oriented productivity solutions for sharing and integrating effort and data under time and other resource constraints.

5. **Ethics.** Students will understand the imperatives underlying research and data ethics and will scope out the societal context and implications of their work by applying conceptual frameworks from the humanities and social sciences – to identify ethical, legal, and social issues surrounding data collection and analysis; to creatively develop and evaluate responses to these issues; and to implement appropriate responses.

6. **Applications.** Students will apply informatics theory and data-management methods to address boundary-spanning problems pertaining to business, management, economics, sociology, psychology, cognitive science, environmental science and engineering.

How the course requirements ensure program learning outcomes will be met

The lower and upper division course requirements outlined in subsection 2.1 support the 6 PLOs outlined in previous subsection.

- PLO 1 – *Methods* will be covered in lower-division core & DSA Foundations courses, including DSA 001, CSE 015, ECON 010, and DSA 100, DSA 1X1, 1X2 and COGS 125.
- PLO 2 – *Communication* will be covered by numerous project-based courses drawing on technical and graphical communication, in particular MIST 135, DSA-1X2 and DSA 120/121 (Capstone).

- PLO 3 – *Design* will be covered by several advanced courses in which students must work within real-world constraints to develop feasible solutions. In particular, ECON 110, MIST 130, DSA 1X1, DSA 1X2, and DSA 120/121 (Capstone).
- PLO 4 – *Team* will be covered by DSA 001, MIST 050, and DSA 120/121.
- PLO 5 – *Ethics* will be covered by PHIL 002, MIST 050, DSA 1X2 and COGS/PHIL/DSA 1X3.
- PLO 6 – *Applications* will be covered by, DSA 1X1, DSA 1X2, DSA 120/121 and DSA PPP Electives.

2.4 Assessment

We plan to follow standard UCM protocol for administering program assessment. Internal and external reviews will be obtained and administered by relevant CIS/EBM/MCS/POLISCI faculty, in consultation with faculty from Math and CSE departments who offer the lower-division DSA Core courses. As such, lower-division program assessment will mirror other STEM programs that have similar 1st and 2nd year quantitative foundations courses. Upper-division program assessment will focus on: (a) incoming and outgoing transfer rates, in particular with respect to the CSE undergraduate major; (b) suitability of DSA 001 as primer courses for students with little or no computing background prior to enrolling in the program; (c) program/faculty quality and support; (d) student educational outcomes and achievements; (e) capstone outcomes.

2.5 Minimum and maximum number of credits allowed for major

Range: 68-80 (Minimum assumes a student testing out of Math 011, 012 and Econ 001 based on sufficient AP test score; Maximum corresponds to 20 required courses)

2.6 How student can accomplish major (including pre-requisites) in 4 years

The lower-division course requirements are similar to other STEM majors in that students must pass fundamental math courses (Calculus I and II) to advance to upper division courses. Students who fail to pass these courses will have difficulty in completing the major in 4 years, but this would also be the case in many STEM majors. Students must also complete DSA 001 and 100, ideally upon their first attempt, to stay on track. Once students complete these lower-division courses that are common pre-requisites for other upper-division courses, there is additional flexibility built into the program by way of allowing students to choose 6 upper division electives, which provide ample room for student modification and course correction. The remaining challenge for students to complete in 4 years is to complete the DSA Capstone I/II on the first attempt.

2.7 How transfer students can complete the major in 2 years

A substantial number of first and second-year courses can be completed via transfer credits. The principal challenge would be to satisfy the requirements of DSA 001, which are unique to the campus. One other lower-division courses that would be difficult to satisfy with transfer credits is MIST 050 Entrepreneurship. Nevertheless, a strong dedicated student could likely test out of DSA 001, and then take MIST 050 in Year 3 or 4.

3. Resource requirements

3.1 Faculty/instructional requirements

This major draws the majority of its courses from existing classes across the university, in particular the standard compulsory introductory CSE and Mathematics courses common to STEM majors across campus, such as introductory Calculus. Only one new required course (DSA 001) will need to be developed before the program can launch, and it is being piloted as MIST 190 in Fall 2021; DSA 100 would be needed by the second year; and DSA 120 and 121 would be need by the fourth year of the first cohort.

Every fall (spring), one faculty member—preferably but not necessarily from the CIS/EBM/MCS/POLISCI department—will teach Foundations of DSA 001. DSA 100 is a computationally-oriented introduction to essential concepts of statistics, probability, linear algebra, and differential equations tailored around a data- and graphical-oriented introduction to these fundamental topics, as opposed to the formulaic introduction found in mathematics courses which takes the traditional theoretical route including demonstration by proof. And Capstone I and II (DSA 120/121) are standard team-based laboratory courses compulsory for seniors. Other new DSA electives will be developed as needed, e.g. in methods of advanced visualization, machine learning and natural language processing. Hence, the minimal additional resource needs to staff five new courses will be split among three existing departments.

Teaching assistants for the five new compulsory courses will be draw from CIS/EBM/MCS/POLISCI graduate programs, or possibly from other STEM graduate programs.

The program will adopt the School of Engineering Policy on Laptops (effective Aug. 2019), such that students will be expected to have their own personal laptop suitable for computational analysis (e.g. minimum 16 GB RAM which has become the standard for Apple laptops). The program will provide loaner laptops to support students facing financial burdens. For this reason, any classroom with sufficient number of power outlets will suffice, and so no specialized computing resources will be needed for laboratory sections.

Advanced courses may request educational access to the MERCED Computing Cluster to showcase its high-performance computing capabilities. The program will integrate with short courses and other computational training offered by the Office of Information Technology and the Library, such as the Software, Data and Library Carpentry workshops which cover concepts such as Data Literacy, Collections, Data Services, and GIS.

Regarding other Library/Collections needs, this program does not anticipate any, as it will leverage open data sources and existing open textbook collections. A principle example of the latter is the widely celebrated O'Reilly/Safari repository of data science textbooks, available at no student cost through the UC's federated agreement, and accessible to students and faculty via standard UCM single sign-on.

3.2 Needs for staff or instructor FTE

No additional staff or instructors are needed to launch the major.

4. Potential for non-majors to participate

All courses except for DSA 120/121 (Capstone) will be open to non-majors satisfying prerequisites. Math majors completing the Computational and Data Sciences emphasis track will be considered as exceptions for these capstone courses.

We will also design DSA 001 Foundations of DSA in such a way that it satisfies the GE Quantitative Reasoning and Language badge. These two courses will be offered in sequence every Fall and Spring semester, providing new options for undergraduates seeking to fulfill these GE requirements.

5. Timeline for Implementation

5.1 Timeline

First cohort in Fall 2023 – Freshman only.

Sophomore (Junior) transfer admissions considered in Fall 2024 (Fall 2025).

The main short-term considerations for rolling out the program are certifying the freshman Foundations (DSA 001) and second-year Intro Quantitative Analytics (DSA 100) courses. Given that these courses will establish core knowledge and computing frameworks integrated across the remainder of the degree, these courses would benefit from being synergistically developed, in principle by a small team of the program's core faculty. Two capstone courses (DSA 120 and 121) will not need to be taught until three years after commencing. As such, given the relatively small resource requirements of the program, we believe the program could feasibly launch in Fall 2022.

Long-term considerations regarding the successful launch of the program include developing a sufficient number of local industry and public partners who are willing to sponsor data-oriented capstone projects, which may involve special data privacy agreements. To address this gap in the short term, faculty and other entities on campus (e.g. the Library or Facilities Management) will be able to develop suitable projects for capstone teams.

5.2 Outline of teaching plan for required courses

Major Preparation courses

MATH 011, MATH 012, CSE 015 and CSE 019: are offered F/SP annually by faculty from the MATH and CSE departments.

ECON 001 and COGS 001: offered F/SP annually by faculty from EBM and CIS

PHIL 002 and PHIL 003: offered F/SP annually by faculty from CIS

Lower-Division Core courses (non-electives)

ECON 001 and COGS 001: offered F/SP annually by faculty from EBM and CIS

DSA 001: Foundations of DSA – A. Petersen (MCS)

MIST 050: Entrepreneurship – R. McBride (MCS)

ECON 010: Statistical Inference – Taught regularly in EBM department by Lecturer

Upper-Division courses (non-electives)

DSA 100: Intro to Quantitative Analytics – TBD (MCS/CIS)

ECON 110: Econometrics – A. Johnston (EBM)

COGS/PHIL 1XX: Data Ethics – H. Gunn (CIS)

MIST 130: Stat. Data Anal. and Opt. in R for Decision Support – L Westerling or S. Castro (MCS)

MIST 134: Methods of Data & Network Science – A. Petersen (MCS)

DSA 1X1: Machine Learning & Natural Language Processing – TBD (MCS/CIS/EBM)

DSA 1X2: Interactive Data Visualization – L. Padilla (CIS)

DSA 120 & DSA 121: Capstone 1 & 2 – Staff (MCS/CIS/EBM)

6. Governance

The proposed DSA major program is being submitted for review by MCS through the School of Engineering. However, the DSA major is intended to be offered jointly by the four departments – CIS, EBM, MCS, and POLISCI – comprising the future Gallo School of Management Once the new school is established, the four contributing departments will define a governance model whereby a chair, rotating equitably across participating faculty departments, will lead an executive committee (comprised of a single representative from each participating department) to address program development, instructor and TA assignments, and other needs. This committee will also be tasked with undergraduate program assessment according to the standard campus procedures. This committee will also manage the annual teaching plan in coordination with relevant department chairs, as the core curriculum spans multiple departments.

Course List

Required Major Preparation [24 Units]

All Data Science and Analytics majors are required to complete the following lower-division preparation courses.

Mathematics and Computer Science Requirement [16 units]

Complete the following course:

- CSE 015: Discrete Mathematics Units: 4
- CSE 019: Introduction to Computing Units: 4
- MATH 011: Calculus I Units: 4
- MATH 012: Calculus II Units: 4

Foundation Requirement [8 units]

Complete the following courses:

- ECON 001: Introduction to Economics or COGS 001: Introductions to Cognitive Science
- PHIL 002: Introduction to Ethics or PHIL 003: Contemporary Moral Problems

---

Lower Division Major Requirements [12 Units]

Data Science and Analytics Lower Division Core Requirement [12 units]
Complete the following courses:

- DSA 001: Units 4
- ECON 010: Statistical Inference Units: 4
- MIST 050: Introduction to Entrepreneurship Units: 4

Upper Division Major Requirements [44 Units]

Upper Division Data Science and Analytics Principles Requirement [24 units]
Complete the following courses:

- DSA 100: Introduction to Quantitative Analytics Units: 4
- ECON 110: Econometrics Units: 4
- MIST 135: Technical Communications and Visualization Units: 4
- MIST 130: Statistical Data Analysis and Optimization in R for Decision Support Units: 4 or MIST 134: Methods of Data and Network Science Units: 4
- COGS XXX or DSA 1X3: Units 4
- DSA 1X1 or DSA 1X2: Units: 4

Data Science and Analytics Capstone Requirement [4 units]
Complete the following two courses:

- DSA 120 Units: 2
- DSA 121 Units: 2

Application Domain Elective [16 units]
Complete at four additional courses in from the following application domain course list.

The Environment and Sustainability

- ESS 110: Hydrology and Climate
- ESS 113: Sustainability in the Anthropocene
- ESS 132: Applied Climatology
- ESS 141: Environmental Science and Policy
- MIST 132: Geographic Information Systems Analysis in Management
- PHIL 122: Bioethics

Policy and Decision Making

- ECON 120: Economics of the Environment and Public Policy
- ECON 153: Judgement and Decision Making
- MIST 131: Data Governance for Analytics Projects
- POLI 112: Public Policy: Analysis, Strategy and Impact
- POLI 120: Voting Behavior, Campaigns and Elections
- POLI 170: Theoretical Models of Politics
- POLI 175: Advanced Analysis of Political Data
- PHIL 108: Political Philosophy

Understanding and Modeling People

- CSE 175/COGS125: Intro to Artificial Intelligence
- COGS 103: Intro to Neural Networks in Cognitive Science
- COGS 104: Complex Adaptive Systems
- COGS 105 Research Methods for Cognitive Science
- COGS 122: Modeling Social Behavior
- COGS 128: Cognitive Engineering
- COGS 130: Cognitive Neuroscience
- COGS 170: Judgement and Decision Making

Other Elective Courses

- ENVE 155: Decision Analysis in Management
- MIST 130: Statistical Data Analysis and Optimization in R for Decision Support
- MIST 134: Methods of Data and Network Sciences
- MIST 170: Information Systems for Management
- MIST 1XX: Design Thinking
- MIST 1XX: User Experience
- MIST 1XX: Data Analysis in Python, JupyterLab, & Plotly
- MIST 190: Special Topics Course

Example 4-year Student Course Calendar

Year 1: Fall - 16 units	Year 1: Spring - 16 units	Year 2: Fall - 16 units	Year 2: Spring - 16 units
Math 011: Calculus I	Math 012: Calculus II	MIST 050: Intro to Entrepreneurship	CSE 015: Discrete Mathematics
CSE 019: Intro to Computing	ECON 001 / COGS 001	*DSA 100: Intro to Quantitative Analytics	ECON 010: Statistical Inference
*DSA 001: Foundations of DSA	WRI 010: College Reading and Composition	PPP Elective	PPP Elective
SPRK 001: Spark Seminar	PHIL 002 / PHIL 003	GE/Elective	GE/Elective

Year 3: Fall - 16 units	Year 3: Spring - 12 units	Year 4: Fall - 14 units	Year 4: Spring - 14 units
ECON 110: Econometrics	*COGS/PHIL/DSA 1X3	*DSA 1X1 or DSA 1X2	MIST 130 / MIST 134: Advanced Methods Elective
MIST135: Technical Communications and Visualization Skills or COGS 1X1	PPP Elective	*DSA 120: Capstone I	*DSA 121: Capstone II
GE/Elective	GE/Elective	PPP Elective	GE/Elective
GE/Elective		GE/Elective	GE/Elective

Major Preparation Course
DSA Course Requirement
GE or Elective Course

* New Course