

Contact: Rachel Boon

**REQUEST FOR NEW PROGRAM AT IOWA STATE UNIVERSITY:
BACHELOR OF SCIENCE IN DATA SCIENCE**

Action Requested: Consider approval of the request by Iowa State University to establish a Bachelor of Science in Data Science in the Ivy College of Business.

Executive Summary: The Council of Provosts and Board office support approval of this program.

Background:

Description of proposed program. The proliferation of rich and complex data in science, industry and government is fueling the rapid growth of data science as a discipline. The National Science Foundation Directorate for Mathematical and Physical Sciences Support for the Statistical Sciences—a subcommittee of the Mathematical and Physical Sciences Advisory Committee—states that data science is the “science of planning for, acquisition, management, analysis of, and inference from data”¹. The Bachelor of Science degree program in Data Science is intended to enable students to work in data science. The courses in this major are designed to provide the core skills and problem-solving approaches to compete for leading-edge analytics positions, as well as jobs with significant data science components, e.g., establishing and operating data analysis pipelines², creating significant new components of such pipelines, etc.

This program is intended for students with strong quantitative backgrounds and has the goal of educating students on the technical fundamentals of data sciences, with a focus on developing the knowledge and skills needed to manage and analyze large-scale, heterogeneous data to address a wide range of problems. The B.S. in Data Science consists of

- (1) 38 credit hours in the major core, three credits of which constitute a capstone course that is expected to provide experiential learning;
- (2) Nine credit hours in one of seven elective tracks to examine applications and theory of data sciences in a specific area; and
- (3) 23 credit hours of foundation courses. The capstone course will provide an opportunity for students to apply data science concepts to an application area while working in a multi-disciplinary team setting.

Academic objectives. The program will prepare students with the technical and communication skills to enter the workforce as data scientists. These positions are in high demand today in industry, nonprofit agencies, and government.

Learning Outcomes

After successfully completing the program, students majoring in Data Science will demonstrate

- a. an understanding of and an ability to apply the following data science concepts, tools and methods to data analysis pipelines:

¹ Iain Johnstone and Fred Robert (chairs), “Data Science at NSF”, April 2014.
<https://www.nsf.gov/attachments/130849/public/Stodden-StatsNSF.pdf>

² A data analysis pipeline refers to a conceptual framework, often fully or semi-automated using computational tools, for solving data science problems that consumes data from one end and produces insights from the other end.

- i. Data acquisition: acquire data appropriate for a variety of research designs; understand multiple data contexts; use multiple data storage and representation formats to acquire data from a variety of sources.
 - ii. Data preprocessing: implement appropriate techniques for data preparation, including knowledge of data types (e.g., non-numeric heterogeneous data, categorical data, text, etc.), data schema and data conversion techniques; implement a variety of data cleaning methods, such as working with non-numeric heterogeneous data like dates, categorical data and text data; working with regular expressions and tools like “grep”, “awk”, and dealing with missing data and outliers.
 - iii. Exploratory data analysis: implement best practices and strategies for exploratory data analysis, (e.g., the use of key graphs/charts); design data-driven research questions and perform appropriate descriptive data analyses to answer these questions; implement appropriate descriptive data analysis systems (e.g., data structures and algorithms for computing descriptive statistics) for a given situation.
 - iv. Inferential and Predictive Thinking, Modeling and Analysis: formulate appropriate research questions for a variety of data-driven research studies; create actionable intelligence to facilitate decision making in a data-driven context; determine the source and implications of data dependencies within a research design; implement a variety of statistical models, such as linear regression, generalized linear models and graph-based models; design and conduct predictive data analysis using machine learning tools, test and training techniques; implement predictive data analysis systems (designing and implementing machine learning and statistical inference algorithms); assess and quantify uncertainties in data analysis results.
 - v. Computational Thinking, Data Structures, and Algorithms: develop computational maturity regarding data structures and algorithms; implement data storage and indexing systems, such as distributed and fault tolerant storage; develop and implement scalable algorithms for query processing, such as parallel and distributed data processing; demonstrate efficient use of computational resources, e.g. memory, CPU time, and bandwidth.
- b. an understanding of ethical, legal, societal and economic concerns in application of data science concepts,
 - c. an ability to visualize, interpret and communicate the output of data analysis pipelines to stakeholders, and an ability to function on multi-disciplinary teams using concepts and tools from data science.

Need for proposed program. There is massive demand for data science professionals today both in Iowa and nationally, and this market is expected to grow. Harvard Business Review labels data scientist as the most attractive job of the 21st century³. The Department of Labor projects a 25 percent growth rate in employment for data scientists and analysts through the year 2018. A report from the McKinsey Global Institute forecasts the need for hundreds of thousands of data science jobs in the next decade⁴. Iowa State University is well positioned to fill the need due to its long history in data-driven sciences. Iowa State University’s significant recent investments in the area have served to strengthen the university’s capacity in this field further. The program is designed to complement existing undergraduate degrees and prepare students who have an in-depth understanding of data science in a variety of domains.

³ <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

⁴ McKinsey Global Institute, “*Big Data: The Next Frontier for Innovation, Competition, and Productivity*”, 2011 New York: McKinsey & Co. <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>

Link to institutional strategic plan. Data science is a priority area for Iowa State University. For example, ISU President's Destination 2050 initiative features Big Data as one of the six targets⁵. Iowa State University's Presidential Initiative for Interdisciplinary Research has specifically targeted data driven science⁶. Data science is also one of the strategic areas in several ISU colleges and academic units including LAS⁷.

Relationship to existing programs at the institution. There are no undergraduate programs at ISU that meet the objectives of the Data Science program.

Majors, minors and/or individual courses in various ISU colleges, e.g. Computer Science, Statistics, Business Analytics, Bioinformatics, etc., provide material relevant and useful for those pursuing careers in data science but none provides the breadth of coverage open to a wide range of students as does this proposed program. Some topics covered in the core courses in the Data Science program are not offered at ISU, e.g. principal components of a data science pipeline, data science project management, ethical issues in data science. Some topics covered in this program are covered only in specialized courses available only to individuals within a very small number of majors due to prerequisite requirements, e.g. elements of predictive analysis such as training and test sets; feature extraction; survey of algorithmic machine learning techniques.

In Spring 2017, proposals to establish a Minor and an Undergraduate Certificate in Data Science were approved by the faculty. The minor and the certificate in data science are aimed at providing students in other disciplines with data science skills to enable them to establish and operate data analysis pipeline in their area, while the data science undergraduate major is intended for students with strong quantitative backgrounds and seeks to educate students on the technical fundamentals of data science.

Relationship to existing programs at other colleges and universities. The University of Iowa has a graduate program in business analytics, whereas the current proposal is designed for undergraduate students. The University of Iowa is also currently developing an undergraduate degree in data science that consists of a selected list of courses from the Statistics and Computer Science departments at the university. ISU has reviewed a copy of their degree requirements. The University of Iowa degree proposal is aimed at a select group of students interested in developing deep expertise in Statistics and Computer Science. ISU's proposal targets students who are seeking balanced expertise in Statistics, Computer Science, and applying data science concepts to an application area, in line with ISU's motto of science with practice. The University of Iowa has also introduced a track within their Statistics B.S. degree called "Statistical Computing and Data Science," whereas the current proposal is for a full undergraduate degree. The University of Iowa has also introduced an undergraduate certificate, which started Fall 2015, called "Large Data Analysis Certificate". This certificate includes such courses as database management, high performance computing, optimization techniques and knowledge discovery; whereas the proposed data science major incorporates courses that are designed for data science and focuses on topics such as data acquisition and preparation, exploratory data

⁵ <http://www.destination2050.iastate.edu>

⁶ http://www.vpresearch.iastate.edu/index.cfm/47530/40579/presidential_research_initiative_promotes_big_thinking_in_datadriven_science

⁷ <https://news.las.iastate.edu/2017/01/11/college-of-liberal-arts-and-sciences-designated-as-recipient-of-one-of-iowa-state-universitys-largest-gifts/>

analysis, visualization, predictive analysis, machine learning, etc., that more closely relate to setting up and operating data analysis pipelines.

The University of Northern Iowa currently has no related programs.

Luther College, a private liberal arts college in Iowa, also created a Data Science Major and a Minor degree in 2016. Compared to the proposed program, Luther's Data Science program has much less emphasis on computational thinking, e.g. algorithm design, and inferential thinking concepts. Furthermore, this ISU proposed program incorporates integrated courses that are designed from the ground up for the Data Science discipline. In addition to these fundamental differences between these programs, Luther is targeting a different set of students.

Unique features. An important part of the statistical center at ISU is the Statistical Laboratory, which was established in 1933 as the first statistics unit in the US. In recent years, Iowa State University has placed great emphasis on developing campus-wide expertise in big data and data science, and ISU is an active member in the Midwest Big Data Hub. As part of a recent presidential initiative, Iowa State hired 20 faculty members during the 2013-2014 academic year, including 10 in the area of big data⁸.

Student demand. While student demand already exists as evidenced by creation of similar programs across the United States⁹, further demand for incoming students at ISU will be developed by marketing the program and the job opportunities available to Data Science majors. The Business Analytics Master's degree offered by the ISU Ivy College of Business since Fall 2015 has been able to attract yearly cohorts of about 40 graduate students. The number of undergraduate students interested in Data Science should be at least at this level, if not higher (probably closer to the numbers seen by the Computer Science program).

Workforce need/demand. Data Science is a new field, such that estimating the number of students is challenging, but the demand from companies for qualified data scientists is huge. Companies are actively trying to recruit data scientists using even unconventional methods (e.g. the Kaggle job board at <https://www.kaggle.com/jobs>). The median salary for a data scientist is reported as \$123,461 (as of January 2018; <http://www1.salary.com/Data-Scientist-Salaries.html>), with a range of \$107k - \$138k.

Resources to establish a high-quality program. Current facilities and equipment are adequate to establish and maintain a high quality program.

Faculty positions for teaching the existing elective courses listed in the appendix will continue to be needed to sustain this data science program. As enrollment expands, new resources or reallocation of available resources will be necessary. The College of Liberal Arts and Sciences (LAS) will commit resources from within the college that will be needed to propose, develop, and teach the new data science courses listed in Appendix A (see 13.). To the extent that faculty members outside of LAS are involved in teaching the four core courses, other colleges will provide the resources to support the cost of the existing faculty salaries.

Existing facilities associated with existing courses, e.g., data science electives will continue to be used. As the program expands, there may be a need for new computer/statistics lab facilities. The Iowa State University Computation Advisory Committee (CAC) has the responsibility of

⁸ <http://www.provost.iastate.edu/faculty-and-staff-resources/hiring/opportunities>

⁹ <http://datascience.community/colleges>

overseeing the expenditure of the Student Technology fee. The program will apply for CAC/LASCAC funds to cover additional expenses associated with computers and undergraduate personnel. No new research or library facilities are needed.

While this program does not have significant supply needs at this time, a small supply budget will be required for various needs, such as printing information brochures to be used at college orientation events to recruit for data science introductory courses.

Financial resources to support the program will come from the College of Liberal Arts and Sciences (college reallocation). In addition, tuition revenue generated from teaching the new data science courses and increased enrollment is anticipated to offset the total costs.

Cost.

	TOTAL COSTS	TOTAL NEW COSTS
Year 1	100,000	100,000
Year 2	253,750	153,750
Year 3	360,556	106,806
Year 4	365,965	5,409
Year 5	421,454	55,489
Year 6	427,776	6,322
Year 7	434,193	6,417

The estimates presented in the table above are based on the following assumptions:

- Year 1 a new tenure-track faculty member will be hired to teach data science related courses. The total new costs reflect average LAS salary and benefits.
- Year 2 a dedicated student service specialist for the Data Science programs would be hired. The total new costs reflect average LAS salary and benefits for a student service specialist as well as misc. cost such as supplies.
- Year 2 and Year 3 one new tenure-track faculty member will be hired each year to help teach lower and upper level data science related courses and to teach multiple sections of lower level courses. A 1.5% yearly inflation in costs is assumed.
- Year 5 a lecturer for the Data Science programs would be hired to offer additional sections.

Projected enrollment.

Undergraduate	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6	Yr 7
Majors	15	30	50	75	95	115	130
Non-Majors	0	0	0	0	0	0	0

Articulation agreements. Any course level articulation agreements with community colleges that are in place at ISU will continue to be honored. Program-level articulation agreements are not planned because in order to finish the Data Science major in four years, it would be essential that students start the program upon matriculation to ISU.

Accreditation. At the time of this writing, accreditation bodies for Data Science programs do not exist. The Data Science major will be included as part of ISU Higher Learning Commission (HLC) accreditation and does not need special HLC approval.

Date of implementation. August 2018

Letters of Support



**COLLEGE OF
LIBERAL ARTS & SCIENCES**

Office of the Dean

240 Schaeffer Hall
Iowa City, Iowa 52242-1409
319-335-2611 Fax 319-335-3755
clas@uiowa.edu www.clas.uiowa.edu

April 5, 2017

Dear Arne Hallam:

Thank you for sending us your proposal for the BS in Data Science to be offered by the College of Liberal Arts and Sciences at Iowa State University.

We are very happy to support this proposal since the field of data science is rapidly growing and is of great importance to almost every business and career path.

We wish you the best as you move forward with your proposal.

Sincerely,

A handwritten signature in black ink that reads "Chaden Afzalali".

Dean of the College of Liberal Arts and Sciences
UI Alumni Association Dean's Chair in the Liberal Arts & Sciences

From: Lang, Joseph B [<mailto:joseph-lang@uiowa.edu>]
Sent: Wednesday, February 22, 2017 11:55 AM
To: Morris, Max D [STAT] <mmorris@iastate.edu>
Subject: UI Data Science B.S., proposal

Hello Max,

It has been too long...I hope all is going well for you and the department.

We are in the process of developing an undergraduate data science major (B.S.) that will be jointly offered by our department (STATS/ACTS) and computer science (CS). This is intended to be a rigorous major that will attract maybe 10-15 highly-motivated students per year. I've attached a draft proposal.

I am interested in your opinion of this proposal. Is there something we are missing, something that is not needed? Do you have anything similar to this offering at ISU?

All the best from Iowa City,
--Joe

P.S. I do hope that we can meet at an upcoming ASA Iowa Chapter meeting...

Joseph B. Lang
Professor and Chair
Department of Statistics and Actuarial Science
University of Iowa, Iowa City, IA, USA
241 SH, +1(1)-319-335-0712
joseph-lang@uiowa.edu
