Dr. Gordon J. Miller
Chair, LAS Curriculum Committee

On behalf of the Data Science Curriculum committee, we are submitting a program proposal for an Undergraduate Major in Data Science for your consideration. The Data Science Curriculum committee consists of the following members from across Iowa State University (ISU) colleges.

1. David G. Acker, Associate Dean, Global Agriculture Programs
2. Raymond W Arritt, Dept. of Agronomy
3. Cameron Campbell, Associate Dean, College of Design
4. Bryan G Clendenen, Dept. of Graphic Design
5. Carolyn J Dill, Dept. of Genetics/Development & Cell Biology
6. Cassandra J Dorius, Dept. of Human Development & Family Studies
7. Shawn Dorius, Dept. of Sociology
8. Arne Hallam, Associate Dean, College of Liberal Arts and Sciences
9. Chinmay Hegde, Dept. of Electrical and Computer Engineering
10. Heike Hofmann, Dept. of Statistics
11. David K Holger, Associate Provost and Dean of Graduate College
12. Matthew Hufford, Dept. of Ecology, Evolution & Organismal Biology
13. Danny J Johnson, Associate Dean, College of Business
14. Wolfgang H Kliemann, Associate Vice President for Research
15. Sigurdur Olafsson, Dept. of Industrial & Manufacturing Systems Engineering
16. Francis Y Owusu, Dept. Chair, Dept. of Community and Regional Planning
17. James M Reecy, Department of Animal Science
18. Sarah M Ryan, Dept. of Industrial & Manufacturing Systems Engineering
19. Sree Nilakanta, Dept. of Supply Chain & Information Systems
20. Hridesh Rajan, Dept. of Computer Science
21. Heather L Rouse, Dept. of Human Development & Family Studies
22. Sriram Sundararajan, Associate Dean, College of Engineering
23. Wallapak Tavanapong, Dept. of Computer Science

The Data Science Curriculum Committee formed a subcommittee and charged it with designing and developing a program proposal for an Undergraduate Major in Data Science. This subcommittee consisted of Pavankumar Aduri, Julie Dickerson, Amy Froelich, Arne Hallam, Heike Hofmann, Hridesh Rajan, Sriram Sundararajan, and Sarah Ryan. The attached proposal is the work of this subcommittee with regular feedback and approval from the data science curriculum committee.

Respectfully submitted,
Hridesh Rajan, Arne Hallam, and Sriram Sundararajan

FORM A
Board of Regents, State of Iowa

REQUEST TO IMPLEMENT A NEW BACCALAUREATE, MASTERS, DOCTORAL, OR FIRST PROFESSIONAL DEGREE PROGRAM
January 28, 2015

THE PURPOSE OF ACADEMIC PROGRAM PLANNING: Planning a new academic degree program provides an opportunity for a Regent university to demonstrate need and demand as well as the university's ability to offer a quality program that is not unnecessarily duplicative of other similar programs offered by colleges and universities in Iowa.

Institution: Iowa State University

CIP Discipline Specialty Title: Data Scientist

CIP Discipline Specialty Number (six digits): 11.0401; 11.0501; 11.0802; 27.0501; 27.00503

Level: B Bachelor's M _____ D _____ FP _____

Title of Proposed Program: Data Science

Degree Abbreviation (e.g., B.S., B.A., M.A., Ph.D.): BS

Approximate date to establish degree: Month August Year 2018

Contact person: (name, telephone, and e-mail) Hridesh Rajan, hridesh@iastate.edu, 515-294-6168
Amy Froelich, amyf@iastate.edu, 515-294-5584

College that will administer new program: Liberal Arts and Sciences

Please provide the following information (use additional pages as needed). Do not use acronyms without defining them.

1. Describe the proposed new degree program, including the following:

    a. A brief description of the program. If this is currently being offered as a track, provide justification for a standalone 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 program is intended for students at Iowa State University with the goal of enabling them to work in data science. The courses in this data science major are designed to provide students with 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 pipelines[2], creating significant new components of such pipelines, etc.

This Bachelor's of Science degree program in Data Science 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 credits hours in the major core, three credits of which constitute a capstone course that is expected to provide experiential learning; (2) 9 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.

    b. A statement of 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:
    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,

[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.

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

d. an ability to function on multi-disciplinary teams using concepts and tools from data science.

Assessment

To assess, in part, that the learning outcomes for the data science major are achieved, the proposed program will use the following new courses created for the data science program as well as other courses as shown in Table 1.

- DS 201 – Introduction to Data Science
- DS 202 – Data Acquisition and Exploratory Data Analysis
- DS 303X – Concepts and Applications of Machine Learning
- STAT 347X – Probability and Statistical Theory for Data Science
- DS 401 – Data Science Capstone

Brief descriptions of these courses are attached to this program proposal. Two of these courses (201 and 202), as well as the capstone course DS 401, were submitted to the College of Liberal Arts and Sciences (LAS) Curriculum Committee in Spring 2017 (all of these courses were submitted with the LAS designator: 201XX, 202X, and 401X), and approved. The course

proposals for DS 303X and STAT 347X will be submitted in Fall 2017.

Table 1 Course Assessment Plan

Outcome Area	Sub-Area	Course Number and Name
a.	i. Data Acquisition	DS 202 – Data Acquisition and Exploratory Data Analysis COM S 363 – Introduction to Database Management Systems
	ii. Data Preprocessing	DS 201 – Introduction to Data Science DS 202 – Data Acquisition and Exploratory Data Analysis COM S 363 – Introduction to Database Management Systems CPR E 419 – Software Tools for Large-Scale Data Analysis
	iii. Exploratory Data Analysis	DS 202 – Data Acquisition and Exploratory Data Analysis STAT 301 – Intermediate Statistical Concepts and Methods STAT 457 – Applied Categorical Data Analysis COM S 228 – Introduction to Data Structures
	iv. Inferential and Predictive Thinking, Modeling and Analysis	DS 303X – Concepts and Applications of Machine Learning STAT 301 – Intermediate Statistical Concepts and Methods STAT 347X – Probability and Statistical Theory for Data Science STAT 457 – Applied Categorical Data Analysis
	v. Computational Thinking, Data Structures, and Algorithms	COM S 228 – Introduction to Data Structures COM S 230 or CPR E 311 – Discrete Computational Structures/Theoretical Foundations of Computer Engineering COM S 311 – Design and Analysis of Algorithms CPR E 419 – Software Tools for Large-Scale Data Analysis
b.		DS 201 – Introduction to Data Science DS 303X – Concepts and Applications of Machine Learning
c.		DS 202 – Data Acquisition and Exploratory Data Analysis
d.		DS 303X – Concepts and Applications of Machine Learning DS 401 – Data Science Capstone

c. What the need for the program is and how the need for the program was determined;

There is a 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 century[3]. 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 decade[4]. 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. Understanding of the domain and context of data science problems is considered a critical skill for a data scientist and this program's design recognizes this.

d. The relationship of the proposed new program to the institutional mission and

[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

how the program fits into the institution's and college's 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 targets[5]. Iowa State University's Presidential Initiative for Interdisciplinary Research has specifically targeted data driven science[6]. Data science is also one of the strategic areas in several ISU colleges and academic units including LAS[7].

e. The relationship of the proposed new program to other existing programs at the institution; describe how the proposed program will enhance other programs at the university. Will the proposed program duplicate existing programs at the university?

There are no other undergraduate programs at ISU that meet the overall objectives of the Data Science program. In Spring, 2017, proposals to establish a Minor and an Undergraduate Certificate in Data Science were submitted and approved up to the ISU Faculty Senate Curriculum Committee (as of this writing, the proposals still needed to be read and voted on by the ISU Faculty Senate).

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.

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 has the goal of educating students on the technical fundamentals of data science.

f. The relationship of the proposed new program to existing programs at other colleges and universities in Iowa, including how the proposed program is different or has a different emphasis than the existing programs.

As of this writing, there are no such programs at the Regent's universities in Iowa. 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 carefully selected list of courses from the Statistics and Computer Science courses at the university, and we have received a copy of their degree requirements. While the University of Iowa's degree proposal is aimed at a select small group of students interested in developing deep expertise in Statistics and Computer Science, ISU's proposal is aimed at an, arguably broader, group of students who are interested

---

[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/

in a balanced expertise in Statistics, Computer Science and 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 an 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 existing 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 analysis, visualization, predictive analysis, machine learning, etc., that more closely relate to setting up and operating data analysis pipelines.

The University of Northern Iowa doesn't currently have any related programs.

Luther College, a private liberal arts college in Iowa, also created a Data Science Major and a Minor degree in 2016. This Data Science Major's core consists of ten courses out of which six are computer science courses, two are statistics courses, and three are subject matter courses (equivalent to "application areas" in this proposal). 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 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.

g. Special features or conditions that make the institution a desirable, unique, or appropriate place to initiate such a degree program.

Iowa State University (ISU) has a longstanding history in data science. 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. For example, 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 data[8]. As mentioned previously, ISU President's Destination 2050 initiative has featured Big Data as one of the six targets[9]. Last but not least, Iowa State University's Presidential Initiative for Interdisciplinary Research has specifically targeted data driven science[10]. In addition to this focus from central administration, Data Science is also one of the strategic areas in several ISU colleges and academic units including the College of Liberal Arts and Sciences[11]. These activities have collectively prepared the university to develop educational programs in Data Science over the last decade.

h. Are the university's personnel, facilities, and equipment adequate to establish and maintain a high quality program?

Current facilities and equipment are adequate to establish and maintain a high quality program. Additional hires (as discussed in items 11 and 13) would need to be made.

---

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

[9] http://www.destination2050.iastate.edu

[10] http://www.vpresearch.iastate.edu/index.cfm/47530/40579/presidential_research_initiative_promotes_big_thinking_in_datadriven_science

[11] 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/

i. How does student demand for the proposed program justify its development?

As stated in 1.c above, there is great demand for data scientists in academia, industry, non-profit sectors, as well as government. For example, a report from the McKinsey Global Institute forecasts the need for hundreds of thousands of data science jobs in the next decade[4]. This program can train students to fill this unmet need. While student demand already exists as evidenced by creation of similar programs across the United States[12], further demand for incoming freshman at Iowa State University will be developed by raising awareness about the program and about the job opportunities available to Data Science majors. Also see item 2. By providing a program devised to meet the demand for data scientists, and promoting the program, a derived demand for admission to the program will result.

2. Describe the state and/or national workforce need and/or demand for graduates of the proposed program currently and in the foreseeable future (provide documentation about the current sources of data used to estimate need and demand).

Data Science is a new field, so 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 $120k (as of February 2017; http://www1.salary.com/Data-Scientist-Salaries.html), with a range of $104k - $135k. The Business Analytics Master's degree offered by the ISU 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).

3. List all other public and private institutions of higher education in Iowa currently operating programs similar to the proposed new degree program. (For comparison purposes, use a broad definitional framework, e.g., such identification should not be limited to programs with the same title, the same degree designation, having the same curriculum emphasis, or purporting to meet exactly the same needs as the proposed program.)

As mentioned previously, there are no similar programs at institutions of higher education in Iowa (see Items 1e and 1f for discussion of programs at the University of Iowa and Luther College), although some peer and regional institutions have instituted degree programs similar to the one proposed:

DIRECTORY OF NON-REGENT IOWA UNIVERSITIES OFFERING DATA SCIENCE DEGREES

Institution	Program Name	Location	Department
Luther College	Data Science	On campus	Computer Science

REGIONAL INSTITUTIONS WITH PROGRAMS

Institution	Program Name	Location	Department/School

[12] http://datascience.community/colleges

Univ. Minnesota-Duluth	Retail Marketing Analytics	On campus	Business & Economics
Winona State University	Data Science	On campus	Mathematics & Statistics
Univ. Nebraska at Omaha	Data Science concentration	On campus	Mathematics
Univ. Wisconsin-River Falls	Data Science and Predictive Analytics	On campus	Business & Economics

PEER INSTITUTIONS WITH SIMILAR PROGRAMS

Institution	Program Name	Location	Department
Auburn University	Data Science	On campus	Business
Arizona State University	Business Data Analytics	On campus	Business
Ohio State University	Data Analytics	On campus	Interdisciplinary
Penn State University	Data Sciences	On campus	Interdisciplinary
Virginia Tech	Computational Modeling and Data Analytics	On campus	College of Science

(Source for tables: http://datascience.community/colleges)

If the same or similar program exists at another public or private institution of higher education in Iowa, respond to the following questions:

a. Could the other institution reasonably accommodate the need for the new program through expansion? Describe collaboration efforts with other institutions.

While no similar programs exist at comparable quality and cost in the State of Iowa (see Items 1e and 1f), another program is being developed at the University of Iowa as of this writing. However, both of these programs could coexist because there is sufficient demand for data scientists (see Items 1c and 2) and strong interest in students at both institutions. Furthermore, as the communication from the University of Iowa indicates (see Appendix B), the University of Iowa program is intentionally being designed to keep enrollments low.

b. With what representatives of these programs has there been consultation in developing the program proposal? Provide a summary of the response of each institution consulted.

An informal e-mail describing our interest in creating an undergraduate major in Data Science along with degree program requirements was sent to both the Computer Science and the Statistics faculty at the University of Iowa and the Computer Science and the Mathematics faculty at the University of Northern Iowa for comments and suggestions on February 24, 2017. The initial informal feedback from both sister institutions has been positive. A formal letter of support from the University of Iowa was received on April 5, 2017. That letter is included.

c. Has the possibility of an inter-institutional program or other cooperative effort been explored? What are the results of this study? (Consider not only the possibility of a formally established inter-institutional program, but

also how special resources at other institutions might be used on a cooperative basis in implementing the proposed program solely at the requesting institution.)

We have not explored the possibility of an inter-institutional program. Undergraduate students generally want to attend a particular institution and as a result, we do not believe there is anything to be gained from an inter-institutional program. However, if a student in our Data Science program takes or has taken a class at either the University of Iowa or the University of Northern Iowa that is a direct substitute for one of the classes required for our program, we will allow that substitution if it meets our university and college policies for such substitutions.

d. Do other colleges in Iowa offer programs similar to the proposed program at comparable quality and cost?

There are no similar programs at comparable quality and cost. The data science program offered by the Luther College is at a substantially higher cost.

e. Are letters of support included with the program proposal?

A letter of support is included from:

- Dean, the College of Liberal Arts and Sciences
- The University of Iowa

Other letters are being collected as of this writing from following.

- The University of Northern Iowa
- Deans of other ISU undergraduate colleges

4. Estimate the number of majors and non-majors students that are projected to be enrolled in the program during the first seven years of the program.

a. Undergraduate

The table below presents anticipated number of majors in the program.

Table 2: Anticipated Number of Data Science Majors

Undergraduate	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6	Yr 7	Yr 8
Freshman	15	20	25	30	35	40	40	40
Sophomore		10	15	20	25	30	35	35
Junior			10	15	20	25	30	35
Senior				10	15	20	25	30
Total	15	30	50	75	95	115	130	140
Non-Majors	0	0	0	0	0	0	0	0

This projection assumes that the program will see an initial enrollment of 15 students, and the number of first year students would increase by five during years 2-6. Subsequently, a steady state enrollment of about 40 new students per year is anticipated. We also estimate that about five students might transition to other programs after the first year, which is a conservative estimate. Beyond year 8, a steady state enrollment of 145 students is anticipated.

b. Graduate

This is an undergraduate program.

c. What are the anticipated sources of these students?

A majority of these students will be new students from the state of Iowa, neighboring states, and international students that are currently not considering Iowa State University as a destination because Iowa State University doesn't have a Data Science B.S. degree. Some students will also transition from open options in various ISU colleges. There is a possibility that a small number of majors will transition from existing ISU undergraduate programs such as Computer Science and Statistics.

5. If there are plans to offer the program away from the campus, briefly describe these plans, including potential sites and possible methods of delivery instruction. Will off-campus delivery require additional **HLC accreditation**?

The program will be offered on campus only.

6. Has the proposed program been reviewed and approved by the appropriate campus committees and authorities?

The approval for the program is in progress.

7. List date the program proposal was submitted to the Iowa Coordinating Council for Post High School Education (ICCPHSE) and results of listserv review.

8. Will the proposed program apply for programmatic accreditation? When?

At the time of this writing, accreditation bodies for Data Science programs have not emerged. The Data Science major will be included as part of Iowa State University Higher Learning Commission (HLC) accreditation in the next cycle. Once the Data Science B.S. degree is well established, we will discuss with our industry partners and with members of our College of Liberal Arts and Sciences advisory councils whether the pursuit of additional accreditations would be worthwhile.

9. Will articulation agreements be developed for the proposed program? With whom?

We will continue to honor any course level articulation agreements with community colleges that are in place at Iowa State University. 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 as a freshman at Iowa State University.

10. Will there be opportunities for student internships?

Given the demand for Data Scientists (see Item 1.c and Item 2), there should be ample opportunities for internships both within Iowa and outside Iowa. The program will leverage LAS career services, and well-established venues such as the LAS Career Fair and the Engineering Career Fair to connect students with potential employers. The existing alumni network of LAS established from such majors as Computer Science and Statistics will also be leveraged.

11. Describe the faculty, facilities, and equipment that will be required for the proposed program.

Faculty members

Faculty positions for teaching the existing elective courses listed in 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 item 13).

Revenue generated from teaching the new data science courses and increased enrollment in other courses is anticipated to offset the expended resources. Resources may come from new revenues or reallocation of existing funds within the relevant units.

To the extent that faculty members outside of LAS are involved in teaching the four core courses, other colleges will need to provide the resources to support those faculty members.

Computers, laboratories, and other facilities

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 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.

Research facilities

No new research facilities are needed.

Library facilities (journals, documents, etc.) in the proposed area

No new resources needed.

Supplies, field work, student recruitment, etc.

While this program doesn't 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. The LAS college will provide these funds.

12. From where will the financial resources for the proposed program come (list all that apply, e.g., department reallocation, college reallocation, grants, new to the university)?

Table 3: Revenue and Costs

SOURCES					TOTAL AMOUNT	
Year	**CUMULATIVE Reallocation**	**Amount**	**Cumulative Expected**	**Amount**	**New Costs**	**Total Costs**
Year 1	College of LAS reallocation	100,000	**Expected** New revenue	0	100,000	100,000
Year 2	College of LAS reallocation	150,000	**Expected** New revenue	103,750	153,750	253,750
Year 3	College of LAS reallocation	150,000	**Expected** New revenue	210,556	106,806	360,556
Year 4	College of LAS reallocation	150,000	**Expected** New revenue	265,965	5,408	365,965
Year 5	College of LAS reallocation	150,000	**Expected** New revenue	271,454	55,489	421,454
Year 6	College of LAS reallocation	150,000	**Expected** New revenue	277,776	6,322	427,776
Year 7	College of LAS reallocation	150,000	**Expected** New revenue	284,193	6,417	434,193

13. Estimate the total costs/total new costs (incremental increases in expenditures) that will be necessary for the next seven years as a result of the new program. Be as specific as possible.

Table 4: Total Costs and Incremental Increases in Expenditures

	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	5,408
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. In Year 1 a new tenure track faculty member will be hired to teach data science related courses such as DS 201 and DS 202. The total new costs reflect average LAS salary and benefits for a new tenure track faculty member. In 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. In 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 such as DS 303 and to teach multiple sections of lower level courses. A 1.5% yearly inflation in costs is assumed. In Year 5 a lecturer for the Data Science programs would be hired to offer additional sections of DS courses. The program is also expected to generate revenues, and a significant portion of these costs would be offset by the new revenue. The $150,000 in Item 12 (see line “Year 7”) is from reallocation of existing funds; the remaining $284,193 is expected to come from increased enrollment.

14. Describe the marketing plan developed to communicate the new program and recruit students.

Data Science is a new and rapidly evolving area at the intersection of many fields, such as mathematics, computer science, computer engineering, information systems and statistics. It is an area that relies on measurable quantities to come to decisions. The ideal student for this program should have a love for numbers and computers, and a curiosity to make sense of problems. Most problems are unstructured and need creative thinking to come up with solutions. Good communication skills are necessary to convey results.

Since Data Science is a new field, the demand from companies for qualified data scientists is huge. The median salary for a data scientist is reported as $120k (as of February 2017; http://www1.salary.com/Data-Scientist-Salaries.html), with a range of $104k - $135k.
The Business Analytics Master’s degree offered by the ISU 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).

Student Recruiting: Besides the above mentioned pull of the discipline itself, the Data Science (DS) major will be administered by LAS college advisers and promoted in the same venues as other LAS majors including but not limited to open option students within LAS and other ISU colleges.

Ease of transition/co-major: the DS undergraduate program is designed such that students with a strong quantitative background should be able to pursue a double major in Data Science without unduly prolonging their studies. In particular, double majors with Statistics or Computer Science are possible in a four-year plan (see attached), double majors with Mathematics and some Engineering degrees are possible within five years.

Application tracks: the DS major is streamlined such that students are exposed to both the practice of data science as well as the theoretical concepts involved. A feature that sets the DS major at ISU apart from other DS programs are the three-course tracks of application areas. Not only does this allow students to pursue Data Science ideas within a particular subject to be better prepared for their future career, it also provides a structured framework for different programs across colleges to get involved with the DS major.

15. Describe the program evaluation plan to determine if the program is meeting the intended objectives, if the expected student enrollment has occurred, funding for the program, and any other components that affect the effective operation of the program.

The program evaluation plan will include evaluation of program outcomes using course grades, student surveys and portfolios at the time of graduation. We will also track graduates' first activities after graduation and advanced degrees earned. Finally, we plan to survey graduates and their employers every five years to obtain their perceptions of the effectiveness of the program in developing data science professionals.

The program evaluation plan for the B.S. program in Data Science will consist of information from coursework, portfolios, surveys, and tracking data. These areas are listed below.

Performance in Foundation courses: For all B.S. graduates, we will track grades in all foundation courses and calculate the overall GPA in this area.

Performance in Courses related to Program Outcomes: For all B.S. graduates, we will track grades and calculate the overall GPA in courses related to the program outcomes, as listed in Item 1b.

Student Portfolios: For all B.S. graduates, we will have student prepare a portfolios of projects from courses in the major. These projects should indicate the depth and breadth of the student's preparation in data science and its applications.

Graduating senior survey of perceived abilities relative to program outcomes: For all B.S. graduates, we will administer a survey at the time of graduation. The survey will be used to gather information on student's perceived abilities relative to the program outcomes.

Students' first activity after graduation: For all B.S. graduates, we will track the first activity after graduation, either acceptance of a position in industry, business or government, etc. or admittance to a masters or doctoral degree program.

Completion of advanced degrees by B.S. graduates: For all B.S. graduates, we will track completion of advanced degrees.

Survey of B.S. graduates: Every 5 years, we will survey B.S. graduates to obtain information about career paths and perceptions of the effectiveness of the program for preparing them for their chosen career. For the first five years of the program we will conduct the survey of B.S. graduates every year instead of every five years.

Survey of employers of B.S. graduates: Every 5 years, we will survey the employers of B.S. graduates to obtain information about their perceptions of the effectiveness of the program in preparing qualified and effective data scientists. Once the program reaches sustained enrollment, an external advisory board would also be created to provide annual feedback on the program curriculum and its execution.

16. Include any additional information that justifies the development of this program.

Appendix A: CURRICULUM

BASIC EDUCATION

ENGL 150	Critical Thinking and Communication	3
ENGL 250	Written, Oral, Visual, and Electronic Composition	3
ENGL 302/314 or ENGL 332/STAT 332	Business Communications/Technical Communications/Visual Communication of Quantitative Information	3
LIB 160	Information Literacy	1
International Perspectives		3
U.S. Diversity (US DIV)		3
Total		*10

GENERAL EDUCATION

Foreign Language	3 years in high school or one year college	0 - 8
Arts and Humanities		12
Natural Science		8
Social Science		9
Total		*29-37

FOUNDATION COURSES

MATH 165	Calculus I	4
MATH 166	Calculus II	4
MATH 207	Matrices and Linear Algebra	3
MATH 265	Calculus III	4
STAT 201	Introduction to Statistical Concepts and Methods	4
COM S 227	Introduction to Object-oriented Programming	4
Total		23

MAJOR CORE

DS 110	Orientation to Data Science	R
DS 201	Introduction to Data Science	3
DS 202	Data Acquisition and Exploratory Data Analysis	3
STAT 347	Probability and Statistical Theory for Data Science	4
STAT 301	Intermediate Statistical Concepts and Methods	4
STAT 457	Applied Categorical Data Analysis	3
COM S 228	Introduction to Data Structures	3
COM S 230 or CPR E 310	Discrete Computational Structures/Theoretical Foundations of Computer Engineering	3
COM S 311	Design and Analysis of Algorithms	3
COM S 363	Introduction to Database Management Systems	3
CPR E 419	Software Tools for Large-Scale Data Analysis	3
DS 303	Concepts and Applications of Machine Learning	3
DS 401	Data Science Capstone	3
Total		38

MAJOR ELECTIVES 9

ELECTIVES 3 - 11

TOTAL 120

B.S. DEGREE in DATA SCIENCE: SAMPLE FOUR-YEAR PLAN

College of Liberal Arts and Sciences
Iowa State University

FRESHMAN YEAR

Semester 1		Semester 2	
ENGL 150	3	STAT 201	4
LIB 160	1	MATH 166	4
DS 110	R	COM S 228	3
MATH 165	4	Arts and Humanities Choice	3
COM S 227	4		14
Social Science Choice	3		
	15		

SOPHOMORE YEAR

Semester 3		Semester 4	
ENGL 250	3	DS 202	3
DS 201	3	STAT 301	4
COM S 230 or CPR E 310	3	MATH 207	3
MATH 265	4	Social Science Choice	3
Natural Science Choice	4	Arts and Humanities Choice	3
	17		16

JUNIOR YEAR

Semester 5		Semester 6	
ENGL 302/314/322	3	DS 303	3
STAT 3XX	4	STAT 457	3
COM S 311	3	CPR E 419	3
COM S 363	3	Arts and Humanities Choice	3
Elective/Foreign Language	3-4	Elective/Foreign Language	3-4
	16-17		16-17

SENIOR YEAR

Semester 7		Semester 8	
Major Elective	3	DS 401	3
Major Elective	3	Major Elective	3
Arts and Humanities Choice	3	Social Science Choice	3
Natural Science Choice	4	Electives	6
	13		15

Students in all ISU majors must complete a three-credit course in U.S. diversity and a three-credit course in international perspectives. Check (http://www.registrar.iastate.edu/students/div-ip-guide/usdiversity-courses or http://www.registrar.iastate.edu/students/div-ip-guide/IntlPerspectives-current) for a list of approved courses. Discuss with your adviser how the two courses that you select can be applied to your graduation plan.

LAS majors require a minimum of 120 credits, **including a minimum of 45 credits at the 300/400 level**.

You must also complete the LAS foreign language requirement.

Description of New Data Science Courses

DS 201: Introduction to Data Science. Cr. 2-2. F.S.
(Approved by LAS Curriculum Committee Spring, 2017)
Prerequisites: *1-½ years of high school algebra*

Description:
Data Science concepts and their applications; domain case studies with applications in various fields; overview of data analysis; major components of data analysis pipelines; computing concepts for data science; descriptive data analysis; hands-on data analysis experience; communicating findings to stakeholders, and ethical issues in data science.

Course outcomes/objective:
After completing this course, students should be able to:
1. Define key ideas in data science and data analysis
2. Identify major components and various phases of a data science pipeline
3. Construct programs in a high-level language to solve basic data science problems
4. Identify major components of descriptive data analysis
5. Define key ideas in data science project management and identify ethical issues in data science

Course content/major topics to be addressed:
Major topics include:
- Introduction – What is Data Science? Motivating case studies.
- Overview of Data Analysis
    - Major components of a data analysis pipelines
    - Exploration of the data science process
- Computing for data science
    - data types
    - operations
    - control structures
    - functions
    - scoping rules
    - iterations
    - running and debugging programs.
- Descriptive Data Analysis
    - The fundamental concepts and methods of statistics
    - How to interpret graphical and numerical summaries of data
    - Reason and assumptions behind the calculations
    - The correct interpretation of results
- Hands-on data analysis experience
- Communication findings to stakeholders
    - Written, oral, verbal and electronic forms of communication
- Ethical issues in data science

DS 202: Data Acquisition and Exploratory Data Analysis. Cr. 2-2. F.S.

(Approved by LAS Curriculum Committee Spring, 2017)

Prerequisites: *DS 201*

Description:

Data acquisition: file structures, web-scraping, database access; ethical aspects of data acquisition; types of data displays; numerical and visual summaries of data; pipelines for data analysis: filtering, transformation, aggregation, visualization and (simple) modeling; good practices of displaying data; data exploration cycle; graphics as tools of data exploration; strategies and techniques for data visualizations; basics of reproducibility and repeatability; web-based interactive applets for visual presentation of data and results. Programming exercises.

Course outcomes/objective:

After completing this course, students should be able to:

1. Define key file storage structures and methods to acquire data from those structures
2. Identify different types of data displays and the purpose for the same
3. Construct a data analysis pipeline in a high-level language to solve basic data science problem involving steps such as filtering, transformation, aggregation, visualization
4. Apply strategies and techniques for data visualizations
5. Define key ideas in reproducibility and repeatability and identify ethical issues in data science

Course content/major topics to be addressed:

Major topics include:

- Introduction – What is Data Acquisition? What is Exploratory Data Analysis? Motivating case studies.
- Data acquisition
    - file structures, web-scraping, and database access
    - ethical aspects of data acquisition
- Types of data displays
- Numerical and visual summaries of data
- Pipelines for data analysis
    - filtering, transformation, aggregation
    - visualization and (simple) modeling
- Best practices for displaying data
- Data exploration cycle
- Graphics as tools of data exploration
- Strategies and techniques for data visualizations
- Hands-on data analysis experience
- Communicate findings to stakeholders
    - Written, oral, verbal and electronic forms of communication
- Basics of reproducibility and repeatability

DS 303X: Concepts and Applications of Machine Learning. Cr. 3-0. F.S.

(To be submitted, Fall, 2017)

Prerequisites: *STAT 347*

Description:

Machine learning concepts such as training and test sets; feature extraction; principles of machine learning techniques; regression; classical pattern recognition methods; advanced topics in pattern recognition; unsupervised learning techniques; assessment and diagnostics: overfitting, error rates, residual analysis, model assumptions checking, feature selection; communicating findings to stakeholders in written, oral, verbal and electronic form, and ethical issues in data science.

Course outcomes/objective:

Upon successful completion of the course, students will have an understanding of the key concepts of machine learning and learn a variety of machine learning-related concepts such as linear models, nonparametric regression, logistic regression, and support vector machines. Students should be able to:

1. Define key concepts in machine learning such as training and test sets, and feature extraction
2. Apply machine learning techniques to solve data science problems
3. Evaluate and diagnose learned data models
4. Understand ethical concerns in data science
5. Communicate the output of data analysis pipelines to stakeholders
6. Function on multi-disciplinary teams

Course content/major topics to be addressed:

Major topics include:

- Introduction – What is machine learning? Motivating case studies.
    - Basic machine learning concepts: training and test sets, feature extraction
- Principles of machine learning techniques
    - Regression
        - Ordinary least squares
        - Inference for linear regression
        - Robustness
        - Beyond linearity: nonparametric regression, splines, local polynomial regression
    - Classical pattern recognition methods and their assumptions
        - Logistic regression
        - Bayes theorem (LDA, QDA, Naive Bayes)
        - K-nearest neighbors
    - Advanced topics in pattern recognition
        - Support vector machines and its variants
        - Tree based methods (Bagging, Random forest, Boosting)
    - Unsupervised learning
        - Principal component analysis
        - Clustering (k-means, hierarchical)
- Model assessment and diagnostics
    - overfitting, error rates, residual analysis, model assumptions checking
    - introduction to feature selection
- Selected applications in various domains
- Ethical issues in data science

STAT 347X: Probability and Statistical Theory for Data Science. Cr. 4-0. F.
(To be submitted, Fall, 2017)
Prerequisites: *MATH 207, MATH 265, STAT 301*

Description:
Introduction to probability; distribution functions and their properties; classical discrete and continuous distribution; sampling distributions; theory of estimation; theory of inference; use of R statistical package for simulation and data analysis.

Course outcomes/objective:
After completing this course, students should be able to:

1. Describe sample spaces and apply appropriate probability rules.
2. Use appropriate counting methods, such as combinations and permutations.
3. Apply independence of events to calculate the probability of an event; compute conditional probability; and apply Bayes theorem.
4. Calculate probabilities and moments (mean, variance, etc.) for discrete and continuous random variables, and develop and apply discrete and continuous distributions for a given context
5. Describe and simulate a sampling distribution for the mean and variance.
6. Apply the Central Limit Theorem to the sampling distribution of the mean.
7. Describe the relationship between the normal, t, chi-square, and F distributions and apply to common methodology.
8. Determine the Maximum Likelihood Estimator (MLE) for a given model using calculation or simulation, and contrast MLE with other methods of estimation.
9. Evaluate different estimators of the same parameter using MSE, bias-variance tradeoff, optimality, sufficiency, UMVUE.
10. Develop Hypothesis Tests for parameters using principles from Uniformly Most Powerful Tests and Likelihood Ratio Tests, develop confidence intervals for parameters
11. Use simulation to motivate understanding of concepts, connect theoretical results to methodology, and use data analysis and simulation in the analysis of methodology

Course content/major topics to be addressed:
Major topics include:

- Introduction to Probability: sample spaces and probability rules, counting methods, independence, conditional Probability, Bayes Theorem
- Discrete Distributions: General Distribution – development, calculating probabilities and moments, Classical Distributions – Bernoulli, Binomial, Poisson, Hypergeometric, negative binomial
- Continuous Distributions: General Distribution – development, calculating probabilities and moments, Classical Distributions – Uniform, Normal, Exponential, Gamma, Beta
- Sampling Distributions: Mean, variance, Central Limit Theorem, Relationship between normal, t, chi-square and F distributions
- Theory of Estimation: Maximum Likelihood Estimation (contrast with Method of Moments and Least Squares), Properties of Estimators - MSE, bias-variance tradeoff, optimality, sufficiency, UMVUE
- Theory of Inference: hypothesis Tests – Uniformly Most Powerful, Likelihood Ratio, Confidence Intervals
- Simulation and Data Analysis: motivate understanding of concepts, connect theoretical results to methodology, and demonstrate use of simulation in analysis of methodology.

DS 401: Data Science Capstone, Cr. 3-0. F.S.

(Approved by LAS Curriculum Committee Spring, 2017)

Prerequisites: *DS 202; DS 301 or DS 303*

Description:

Students work as individuals and teams to complete the planning, design, and implementation of a significant multi-disciplinary project in data science. Oral and written reports.

Course outcomes/objective:

After completing this course students will have demonstrated in the context of a multi-disciplinary project:

1. an ability to apply data science concepts, tools and technologies to data analysis pipelines,
2. an understanding of ethical, legal, societal, and economic concerns in application of data science concepts,
3. an ability to visualize, interpret and communicate the output of data analysis pipelines to stakeholders, and
4. an ability to function on multi-disciplinary teams using concepts and tools from data science.

Course content/major topics to be addressed:

Students will conduct data science projects and learn about data science topics such as ethical, legal, societal, and economic concerns in application of data science concepts, basics of reproducibility and repeatability

Assessment Plans:

Student performance will be measured via intermediate and final project reports, and oral presentation.

Data Science Application Emphasis Areas

The developments in data science, even though at a nascent stage, have found applications in several areas of science, engineering, humanities, and business. The availability of massive amounts of data combined with the development of new computational and statistical methods are providing unprecedented opportunities for new scientific discovery and is leading into the *fourth paradigm of science* (a phrase coined by Jim Gray), where hypothesis driven discovery is replaced by data-driven discovery. The students will get an opportunity to learn about the impact of data science in diverse scenarios by choosing one the application areas. Students in the Data Science major will choose one area of major electives in order to further their study in the applications and theory of data science. In each area, students will select three courses. The initial proposed areas and courses are:

Big Data: The need to analyze and process massive amounts of data gave rise to new computational problems and models, algorithmic paradigms, and learning algorithms. Students taking this application area will learn these new developments. Students will take three courses from the following courses:

- Com S 435: Algorithms for Large Data Sets: Theory and Practice
- Com S 454: Distributed Systems
- Coms 474: Introduction to Machine Learning
- COMS 424: Introduction to High Performance Computing
- COMS 426: Introduction to Parallel Algorithms and Programming
- COMS 461: Principles and Internals of Database Systems

Engineering Applications: Developments in data mining and machine learning combined with ubiquity of data have found several applications in various engineering disciplines such as fluid dynamics, transportation, cyber physical systems and smart grids. Students will get exposure to data science applications in engineering fields. Students in this application area will take three courses from the following courses:

- EE 425: Data Analytics in Electrical and Computer Engineering
- ME 492: Data Analytics and Machine Learning for Cyber-Physical Systems Applications
- CprE 388: Embedded Systems II: Mobile Platforms
- COMS/CPRE 425: High Performance Computing for Scientific and Engineering Applications

Optimization: Several computational problems that arise in data mining, machine learning, and matrix decompositions can naturally be formulated as various optimization problems such as linear programming, non-linear optimization, convex/concave optimization, and semi-definite programming. Thus optimization plays a critical role in data mining and learning. This application area will expose the students the role of optimization in data science. Students in this application area will take the following three courses

- IE 312: Optimization
- IE 483: Knowledge Discovery and Data Mining
- IE 487X: Big Data Optimization

Security: As data collection mechanisms have become cheap and ubiquitous, massive amounts of highly sensitive and private data have been generated, collected and analyzed by several entities such as financial organizations, search engines, social networks and health care industries. Thus security of such data and protecting the privacy of individuals is a serious challenge. This area introduces the students to fundamentals of privacy and secure in the

context of big data. Students in this application area will take the following three courses:

- COMS 421: Logic for Mathematics and Computer Science
- CprE 431: Basics of Information Systems Security
- Com S 453: Privacy Preserving Algorithms and Data Security
- CPRE 460x: Data-driven Security and Privacy

Software Analytics: Software analytics is becoming a critical component of software development and diagnosis. Software analytics uses data generated during the various phases of software life cycle—software development, testing and deploying; this data includes bug reports, test cases, execution traces/logs, and source code. Software analytics uses techniques and tools from statistics, data mining and predictive analytics. Students in this application area will take three courses from the following courses:

- Com S 342: Principles of Programming Languages
- Com S 413: Foundations and Applications of Program Analysis
- Com S 440: Principles of Compiling
- Com S 474: Introduction to Machine Learning
- CprE 416: Software Evolution and Maintenance

Statistics: Statistical modeling and machine learning plays an important role in data science by "turning data into knowledge." The courses in the Statistics application area will expose students to additional methods not covered in the core. Students will take three courses from the following courses:

- STAT 402: Statistical Design and the Analysis of Experiments
- STAT 407: Methods of Multivariate Analysis
- STAT 421: Survey Sampling Techniques
- COM S 474: Introduction to Machine Learning

Computational Biology: Students taking this application area will apply statistical methods and predictive and exploratory data analysis techniques to mine biological data. Students will take Bio 212 as an Area II – Natural Science General Education requirement and take following three courses:

- BC BIO 322: Introduction to Bioinformatics and Computational Biology
- BC BIO 402: Fundamentals of Systems Biology and Network Science
- BC BIO 444: Bioinformatic Analysis

Appendix B: Communication with Regents Institutions in Iowa

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,

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

THE UNIVERSITY OF IOWA

COLLEGE OF
LIBERAL ARTS & SCIENCES

Department of Statistics & Actuarial Science

241 Schaeffer Hall
Iowa City, Iowa 52242-1409
319-335-0712 *Fax* 319-335-3017
www.stat.uiowa.edu

April 12, 2017

Arne Hallam, Professor of Economics and Associate Dean
College of Liberal Arts and Sciences
237 Catt Hall
Iowa State University
Ames, IA 50011

Dear Dean Hallam:

On behalf of the University of Iowa's Department of Statistics and Actuarial Science, I enthusiastically support the proposal for a Data Science BS degree program to be offered by the College of Liberal Arts and Sciences at Iowa State University.

The past decade has seen an explosion of interest in data science and a commensurate increase in demand for data science professionals. The undergraduate degree you propose will give students the tools needed to land attractive jobs in this growing market. (McKinsey Global Institute forecasts the need for hundreds of thousands of data scientists in the next ten years).

Given the interest and the growing market, there will be ample demand for such data science programs across the state. Indeed, undergraduate data science degree programs should, and will soon, become standard fare at all major universities.

We wish you all the best with your proposal.

Sincerely,

Joseph B. Lang
Professor and Chair
Department of Statistics and Actuarial Science
University of Iowa, 319-335-0712, joseph-lang@uiowa.edu

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

College of Liberal Arts and Sciences
Office of the Dean
202 Catt Hall
Ames, Iowa 50011-1301
515 294-3220
FAX 515 294-1303

DATE: April 10, 2017

TO: Dr. Arne Hallam, Professor of Economics and Associate Dean, College of Liberal Arts and Sciences

FROM: Beate Schmittmann, Dean, College of Liberal Arts and Sciences

SUBJECT: Proposed Bachelor of Science degree in Data Science

The College of Liberal Arts and Science is excited to support the proposal for a Bachelor of Science (B.S.) with a major in data science. According to the National Science Foundation, data science is the "science of planning for, acquisition, management, analysis of, and inference from data"[1]. As detailed in the proposal, the demand for data science professionals is expected to grow explosively over the next decade. ISU has a long record of excellence in fields related to data science, including statistics, computer science, and mathematics. The College of Liberal Arts and Sciences is the academic home of these programs and therefore uniquely suited to offer the new degree.

The college is pleased to provide the resources needed to create and staff the new major as described below and detailed in the program proposal.

- The LAS College will reallocate existing funds ($150,000) to get the program started.
- The LAS College will also allocate new revenue toward program growth, as detailed in Table 3 of the proposal, provided program growth follows or exceeds Table 2. If the program grows more slowly than proposed, resources will be allocated at a lesser rate.
- These existing and expected resources will be used to hire new tenure track faculty, provide dedicated student services support, and add lecturers, as needed (see description below Table 4).
- The LAS college also commits to use existing resources in the Departments of Computer Science and Statistics as well as LAS Student Services to provide advising for Data Science Majors until dedicated student services support is in place.

We look forward to working with our partners from other ISU colleges to help data science majors experience the best possible education and career outcomes.

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

Academic Program Approval Voting Record

This document is to be appended as the last page of the proposal for any new or revised academic program to record the successive votes of approval as the proposal moves through its required review and approval steps. Consult Faculty Handbook Section 10.8 or the Faculty Senate Curriculum Committee website for information regarding Committee review and voting requirements for each action.

Curricular Action: (check appropriate boxes below)

1. X New Program □ Name Change □ Discontinuation □ Concurrent Degree for:
2. X Undergraduate Major □ Graduate Major □ Undergraduate Minor □ Graduate Minor
   □ Undergraduate Certificate □ Graduate Certificate □ Other: _____
3. Name of Proposed Change: Data Science
4. Name of Contact Person: Hridesh Rajan, Amy Froelich, Arne Hallam, and Sriram Sundararajan

   e-mail address: hridesh@iastate.edu, amyf@iastate.edu, ahallam@iastate.edu, and srirams@iastate.edu
5. Primary College: LAS Secondary College: _____
6. Involved Department(s): _____ _____

   _____ _____

Voting record for this curricular action:

Voting Body	Votes			Date of Vote
	For	Against	Abstain	
Data Science Curriculum Subcommittee	5	0	0	3/27/2017
Data Science Curriculum Committee	14	0	0	4/4/2017
LAS Curriculum Committee	7	0	0	4/7/2017
LAS Representative Assembly	18	0	0	4/25/2017
Faculty Senate Curriculum Committee	7	0	0	9/8/2017
Faculty Senate Academic Affairs Council	8	0	0	10/3/2017
Faculty Senate				

[FSCC – November 2013]