26 February 2014

TO: Paul DeLuca, Provost
    Martin Cadwallader, Dean, Graduate School

FROM: John Karl Scholz, Dean

RE: L&S Educational Innovation: Proposal to create a new named option, Master of Science – Statistics, “Data Science”

CC: Steve Ackerman, Associate Dean for the Physical Sciences, Graduate School
    Kelly Haslam, Assistant Dean Academic Assessment and Funding, Graduate School
    Elaine Klein, Assistant Dean for Academic Planning, L&S
    Jocelyn Milner, Associate Provost and Director, Academic Planning and Analysis
    Eric Wilcots, Associate Dean for the Natural and Mathematical Sciences, L&S
    Brian Yandell, Professor and Chair, Statistics

ATTACHMENT: Proposal to create a new Master’s Degree Named Option in Statistics, “Data Science” (February 2014)

On February 4, 2014, the L&S Academic Planning Council considered the attached request to create a new named option in the existing Master of Science – Statistics degree, to be called “Data Science”.

The proposed program is intended to be an educational innovation that will serve students seeking to study Statistics at an advanced level, and to prepare them for modern practice in data science. As you have no doubt observed, this is a rapidly growing area, and there is a growing need in industry and government to train and retrain personnel to address the rapidly expanding quantities of data available.

The proposed program is distinct from the existing stand-alone, theory based MS program which culminates in a labor-intensive master’s exam. This “free-standing” 30-credit master’s program designed to serve students arriving from the workforce with 3-5 years of experience, including pragmatic experience with abstract conceptual thinking and task definition. It may also attract students who are completing undergraduate or graduate programs outside of Statistics, that are “quantitative” studies (Economics, Finance, Mathematics, Engineering) that can be enhanced by skills in data science.
The target enrollment for the program is set to admit 20-40 students per year. Requirements for admission will be consistent with admission to the UW-Madison Graduate School; however, the Department will set a high bar for English language skills (TOEFL score of 93 or better) so as not to strain the English as a Second Language program. Program revenue will defray costs associated with administration of the program, which will be housed in the Statistics Department and which will operate within existing governance and oversight systems therein. The attached document provides detailed information about program requirements, administrative processes, and assessment procedures. L&S and the department are currently developing a memorandum of understanding about the program revenue structure to ensure that revenues will support the faculty and staff who will oversee it. The college and department expect to have an appropriate resource structure and capacity in its courses to allow students to enter the program in Fall 2015.

It deserves mention that the department is working with the Division of Continuing Studies to develop a Visiting International Scholars Program (VISP), which will provide a focused “senior year” experience for students recruited from a selected number of international institutions (all in China for the first year, but broadening thereafter) with which the department has relationships through the faculty. As visiting international students, these undergraduates will be able to study Statistics in courses that can be transferred back to completion of their baccalaureate degrees (just as UW-Madison students are able to do with the courses they take abroad). While here, these visiting students will be exposed to the excellent educational experience available at UW-Madison, and the department believes that some of these students may be excellent candidates who might apply for admission the MS-Statistics “Data Science ” program after completing their baccalaureate degrees. The department expects to pilot the Statistics VISP in Fall 2014.

Together, these programs extend the reach of UW-Madison’s Statistics Department into the workforce and across the world, to meet the growing demand for professionals with deep analytical skills capable of managing, interpreting, and representing the mountains of data generated and needed in our modern world. The expansion into these areas will strengthen existing, traditional programs by bringing new faculty, graduate students, and courses into the curriculum – certainly these changes would happen in time, but the Educational Innovations program provides an opportunity to reach for that change at an accelerated pace, to make a creative and interesting programs like this possible. The L&S Academic Planning Council unanimously endorsed creation of this option, which has my enthusiastic support as well.
A NEW MASTER’S DEGREE PROGRAM OPTION IN STATISTICS
27 February 2014

The Department of Statistics proposes to introduce a new named option in the existing Statistics Master’s Degree program:

Degree name: Master of Science in Statistics
Option: Data Science

The faculty approved this undertaking at a department meeting on October 29, 2013.

1. Overview

We propose to introduce a new option within the Statistics Master’s Degree program designed to prepare students in modern practices for data science grounded by core concepts of statistics. It is designed to meet the demands of a growing need in industry and government for training and retraining to address the rapidly expanding quantities of data available in multiple facets of government, business and society. Students can complete this degree within no more than two years, with many completing within 15 months.

We expect incoming students to have some experience and skill with abstract conceptual thinking and task definition, such as developed in typical engineering, mathematics and computer sciences courses. We expect a level of organization and maturity in logically approaching a complex problem. While we do not have specific ‘tool set’ requirements for the program, we expect some calculus background. As with all analytic endeavors, we use logic and the ability to represent concepts and manipulate symbols in a meaningful manner. We will rely heavily on what amounts to algebraic representations, but we will develop the required tools in context. Thus the students in the program will gain experience representing and analyzing concepts and challenges with arithmetic, algebraic, and algorithmic (programming) tools and expressing the results in plain English as well as with the analytic tools that will be developed. Beyond the technical issues, leadership is needed in envisioning, planning, analyzing, and sharing analytic results from big data.

We propose a MS option designed to train individuals to become leaders in data science. This MS option will ground students in sound data reasoning skills, with a balance of hands-on data discovery, analytics methods, and basic theory of probability and statistics. Advanced skills of programming and analysis of complex data projects will be developed. All training is integrated into the data practicum, in which students enhance their communication and visualization skills, culminating in a masters project designed to simulate a workplace data science experience.

Demand for access to data science is growing rapidly, with over 40 professional programs created in the past two years. The UW-commissioned EAB market study, as well as the Applied Computing market study, point to broad interest in development of a non-traditional program in data science. Based on initial inquiries by other campus collaborators, we expect work-force employee interest from Epic, Johnson Controls, John Deere, GE Healthcare and Boeing. The President's Council of Advisors on Science and Technology (PCAST) reported in 2013 that the US needs one million additional STEM college graduates by 2018, and encourages transforming STEM education to engage students in active learning through evidence-based research experiences (http://eric.ed.gov/?id=ED541511). A 2011 McKinsey report (www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation) projected "a shortage of talent necessary for organizations to take advantage of big data. By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions." That is, the US needs to train roughly 50% more individuals than are
currently employed in such jobs. Many of these new data science staff will emerge through retraining in professional programs such as the one proposed here. We believe we are developing the right focus on technology, communication and leadership to create a highly successful masters in data science.

Some students applying for this masters option may have advanced math preparation, and desire a more theoretical level for core statistical concepts. Students may request approval of this training for part of their credits, but they must take at least 16 credits while in the masters program (in compliance with HLC). These students will advance more quickly through the program and could complete the degree in one year.

Students will begin with training in network skills for remote learners, addressing time management, use of electronic communications, and team communications techniques. These are very important in today’s highly connected world. The next course is an introduction to data discovery. During the year, students will be exposed to modern statistical methods for data analysis and experimental design, as well as core concepts in inference and probability. Electives include advanced statistical computing for big data, data mining and machine learning, spatial/temporal statistics and optional electives such as finance, biostatistics, communications. The program is capped by a data practicum that emphasizes communication skills, teamwork, leadership and in-depth study of a complex data project, ideally drawn from the student’s current or anticipated work setting. The data practicum project serves the place of a Master’s Exam.

Graduates will be comfortable leading big data meetings with top management as well as with other researchers. All students by completion of the program will have a Master’s degree that will serve as a foundation of a career in data science. Some students may choose to pursue further graduate study at high-ranking institutions including, but not limited to, UW-Madison.

The proposed Master’s program will position the UW-Madison Statistics Department as one of a very few top-ranked Departments to offer such a well-rounded, high quality Data Science Master’s option. The influx of talented Master’s students is expected to strengthen our upper-level undergraduate and beginning graduate-level courses. Thus the program will enhance the learning environment for undergraduate Stat Majors and provide more vertical integration from the undergraduate level through MS to the PhD level. Funds generated by the program will be used to strengthen the core teaching and research missions of the Department, for example to support additional teaching assistantships and/or graduate fellowships.

2. Background

The Statistics Department has currently a stand-alone, theory-based Master’s Degree program that involves a labor-intensive Master’s Exam. This Statistics MS has one named option, Biostatistics, which is similar to the basic degree, with stipulation about choice of elective courses. This MS program has been highly successful in industry and government, but is not scalable beyond 24 students per year.

The proposed program option should be attractive to “workforce” employees in industry or government with 3-5 years of work experience, including pragmatic experience with abstract conceptual thinking and task definition. It will also be useful for students completing undergraduate or graduate non-traditional quantitative programs (such as economics, finance, mathematics, or engineering) who want to establish further skills in data science.

3. Name of Programs and Implementation

Degree name: Master of Science in Statistics
Option name: Data Science
Designation on Diploma: Master of Science – Statistics
Designation on Transcript:
4. Current Master of Science Degree vs. Proposed Option

The current Master of Science degree: Students enter the M.S. in Statistics either by application to the Graduate School or by adding a major while in another Ph.D. program (possibly in Statistics). Those in Ph.D. programs of the Statistics Department, as well as other departments at UW Madison, who wish to concurrently earn an M.S. degree in Statistics, will continue to be eligible for the current Master of Science degree (not the named option proposed here). The current degree is aimed to serve students who have other commitments such as TA assignments or enrollment in another graduate level program, and they typically take two to three years to complete the degree. The requirements are 30 graduate level credits (currently 21; being modified to satisfy HLC) in Statistics, with a core of four 600+ level courses that must be passed with a grade of B or higher.

The new named option: In contrast, the new named option is a compressed full-time program with 30 required credits in statistics. These would be completed in no more than two years (at 100%) by students starting at a level considerably lower than PhD-ready students. We expect that many students in the program have sufficient preparation to have some requirements waived based on the completion of equivalent prior work; these students will therefore complete the program in two or possibly three semesters. A detailed description of the curriculum and degree requirements are given in §6 and §7 below.

The new program is designed to prepare students with practical skills for data science, starting from an expectation of modest formal mathematical training. Students admitted to our Ph.D. program in statistics generally already come in with a quantitative background that is superior to that of the students who will be admitted to the new Master’s program. Moreover, the new option is not suitable for students in other UW graduate programs who do not intend to pursue a Ph.D. degree in statistics. These students generally tend to take a specialized selection of courses which enhances the studies in their main subject.

Consideration of students for the department’s Ph.D. program. Admission to the proposed Master’s program will not imply a preferential consideration for admission to the Department’s Ph.D. program. Outstanding students in the M.S. program may be considered for admission to the Ph.D. program, but all such students will be required to submit a formal application to the Ph.D. program.

5. Administrative Structure and Governance

The faculty and the Executive Committee of the Statistics Department will have governance authority over the new Masters option in Data Science (MS-DS). There will be an MS-DS Degree Committee of 3 faculty with rotating membership. Basically, faculty agree to serve for 2-year periods during which they are responsible for admitting a new cohort and monitoring their progress. It is expected that at least one rotating faculty will continue for additional terms. The MS-DS faculty will oversee the admissions process for the new program. The support staff member who serves as MS-DS Coordinator will have added administrative responsibilities; these are similar to the current responsibilities involved in administration of the existing Master’s and Ph.D. programs, but tailored to specific needs of an online program. The MS-DS Degree Committee will be responsible for advising the MS-DS students.

6. Admissions and Degree Requirements

6.1. Admission Requirements

Applicants will have at least 3-5 years of active work experience with responsibilities and
quantitative skills related to big data and/or data science, or have completed another non-traditional program in mathematics or statistics or a related field with a solid mathematical component. For admitted international students who are not native speakers, TOEFL scores of 93 and above are required. Admission should involve in-person or online interviews to assess incoming communication skills.

6.2. Degree Requirements

**Credits:** 30 graduate credits. See Section 7.3 for details.

**Courses:** Courses must be taught through the Statistics Department. Course offered by other departments will be allowed only in exceptional circumstances, and must be approved by the MS-DS Degree Committee.

**Grades:** Students must achieve an overall GPA 3.0 or better in statistics courses.

**Advising:** At admission to the program, students will work with MS-DS faculty to develop a plan of study. Students will meet with a MS-DS faculty advisor early in each summer semester to discuss a plan of study for the academic year.

**Residence requirement:** For resident students, 16 graduate credits must be earned in residence at UW-Madison (as per HLC).

**Time limit:** This is a program designed to be completed in no more than two years. In exceptional circumstances, the Chair of the MS-DS Degree Committee in consultation with the Graduate Program Committee may allow extensions. Students are subject to the Department of Statistics Criteria for Satisfactory Progress.

7. Curriculum

7.1 Current vs. Proposed Curriculum

**Current Curriculum:** Our current Master's degree is reserved for students who apply directly to the program or who are already enrolled in the Ph.D. program in Statistics or in another graduate level program at UW-Madison. There is currently one named options, Biostatistics, with slightly different choice of electives. For details see [http://www.stat.wisc.edu/phd-masters/MS_Degree_Regulations](http://www.stat.wisc.edu/phd-masters/MS_Degree_Regulations). The degree requires 30 (currently 21, to be changed to meet HLC) graduate level credits, all in Statistics, and must pass the Master’s Exam. These traditional options are capacity-limited at 24/year due to the labor-intensive Master’s Exam.

The goal of the current program is to build quantitative and communication skills needed to collaborate in inter-disciplinary research teams, particularly in industry and government. Students learn core methods and theory of statistics, augmented by electives. The “capstone” experience is the Master’s Exam, a one-week marathon in which students are given two “live” (actual, unrehearsed) research problems and must write 12-page reports for each, accessible to the scientist who provided the problem. This real-world experience mimics on-the-job settings, and as such has been highly prized by employers.

**Proposed Curriculum:** In contrast the curriculum in the new named option is designed as advanced training in data science aimed at non-traditional students. Students will either be active in the workforce or come through another non-traditional program. Students are required to earn 30 credits in a wide range of subjects.

The goal of the proposed program is to build complex-data reasoning, visualization and communication skills, and to simulate inter-professional strategic planning experiences that involve making sense of complex data to realize corporate objectives. This MS option will ground students in statistical reasoning, with a balance of hands-on data discovery, analytics methods, and basic theory. Advanced skills of programming and analysis of big data projects will be developed. All training is integrated into the "capstone"
experience of a Data Practicum, in which students must enhance communication and visualization skills, culminating in a masters project that can be used by other professionals, say in a board meeting for strategic planning.

**Proposed Area of Emphasis and Data Practicum:** The degree requirements are sufficiently flexible to allow students to pursue an area of emphasis if desired through electives. The Data Practicum course serves the purpose of a Master's Thesis requirement. This course, being designed in collaboration with the School of Business, will give students experience with a real (or at least realistic) complex data project. Students will need to refine key questions for the project, develop strategies for analysis and visual presentation, and finally present material to a corporate (or corporate-like) audience. That is, students will need an “elevator speech” overview to attract the audience, followed by a tightly developed argument/presentation. The goal is to catalyze strategic planning based on sound analysis of real-world complex data in a corporate setting. The Data Practicum will be designed to be scalable both in terms of student capacity and data magnitude. These data practicum projects may be taken directly from students’ work experience (anonymized as needed).

### 7.2 Other Related Programs

To our awareness, there are no other related programs at UW-Madison. The proposed Masters of Biomedical Informatics is aimed at a health science audience, and is focused on human health. While there conceivably could be some courses taken in common, the thrust of the program is rather different. That program is heavily weighted toward Computer Science courses and concepts as they relate to biomedical investigations.

### 7.3 Proposed Curriculum Detail

The proposed MS courses are below. Courses in **bold** are under development, based largely on existing courses aimed at different audiences. These new courses, being developed with a large Educational Innovation (EI) grant on “Statistics Curriculum Modernization for Data Science Needs”, will be officially approved and ready to teach for Fall 2015.

**MS Curriculum (max 6-9 cr/sem)**

#### Fall #1
- **Stat 627 Data Discovery & Network Skills** (3 cr) [Bates/Gillett/Yandell]
- **Stat 671 Statistical Methods for Analytics I** (3 cr) [Clayton/Zhu/Keuler]
- **Stat 609 Mathematical Statistics I** (3 cr) [Shao]

#### Spring #1
- **Stat 672 Statistical Methods for Analytics II** (3 cr) [Clayton/Zhu/Keuler]
- **Stat 610 Mathematical Statistics II** (3 cr) [Shao]
- *Stat 6xx elective (3 cr)

#### Fall #2
- *Stat 6xx elective (3 cr)
- *Stat 6xx elective (3 cr)
- *Stat 6xx elective (3 cr)

#### Spring #2
- **Stat 6xx Data Practicum** (3 cr) [Yandell/Zhu adapt Stat 998]

*Electives drawn from a range of masters level courses, including:

- **BMI/Stat 6xx Advanced Statistical Programming** (3 cr) [Broman course]
- **Stat 6xx Data Mining** (3 cr) [Loh course; EI sabbatical]
- **Stat/Bus 679 Statistical Finance** (3 cr) [Y Wang/Z Zhang course]
- **Stat 6xx Big Data Analytics Design** (3 cr) [Qian]

**Math stat level depends on student background; workforce professionals may need more pragmatic level (Stat 311-312) while students with strong quantitative backgrounds may be prepared for more mathematically rigorous series (Stat 309-310 or 609-610).**

### 8. Collaborations and Partnerships

We have heard from campus collaborators that leaders at Epic, Johnson Controls, John Deere, GE Healthcare, and Boeing have all expressed interest in training for data science. We are
working with colleagues in the CoE, in particular with EPD via Phil O’Leary, on developing these contacts and a set of advisors. We are also exploring possible collaboration with MATC, Wisconsin Technical Colleges, and the UW System concerning the development of an online program at the state level.

9. Enrollment Projections

The Statistics Department intends to annually admit at least 20 students to this new program, entered as cohorts of 20-40 students, beginning in Fall 2015. The department will monitor the success of the program and accordingly adjust enrollment levels in later years. It is difficult at this time to project the long-term size of this program given the heavy demand for data science expertise in today’s world.

10. Faculty

Faculty members of the Department of Statistics will participate in running the Master’s program, via teaching of courses and advising. Teaching assignments in the Department are rotated among the faculty, and all faculty are expected on average to teach at least one upper-level undergraduate course and one graduate level course each year (out of 3 total per FTE). Thus, many faculty will be teaching the masters level courses designated for the Master’s program. The MS-DS Degree Committee will have 3-4 faculty serving for at least 2-3 years. The MS-DS Degree Committee monitors and maintains appropriate admission standards. Assessment reports will be conducted at least every five years.

11. List of Faculty

The list of Department of Statistics faculty, including academic staff, is available at http://www.stat.wisc.edu/faculty-directory .

12. Financial Support

Students are expected to be self-supporting, with no guaranteed financial support from UW-Madison. The criteria for TA and RA appointments in the department make it unlikely that students in this program would be eligible for support. The required credit load and expected rate of progress render it unlikely that students in this program would have time to take on a graduate assistant appointment.

13. Budget Implications

13.1 Budget Overview

The Master’s program is expected to be fully self-supporting. The Stat Department requests that L&S allow the Department to use some of the profit as Innovative Education funds (exact amount to be determined in consultation with L&S). These funds will be used to strengthen the core teaching and research missions of the Department, for example to support additional teaching assistantships and/or graduate fellowships.

The new program will increase enrollment in our advanced undergraduate and graduate courses. In particular, this increase will require opening new sections in high-enrollment courses and/or additional costs for grading work. For workforce retraining, all courses will ultimately be online, requiring new staffing funded by the program. It will also incentivize modernization of current courses and creation of new courses that are in great demand within the Stat Department and the UW campus community.

The details of the budget will reflect the instructional faculty/staff time devoted to new sections as well as extra students in existing sections. There will also be a budget amount allocated to more hours for grading homework. Since the co-directors of the MS-DS Degree Committee will assume significantly increased responsibility for oversight of all aspects of the Master’s program, summer salary will be provided to reflect the loss of research time during the
academic year. The budget will account for faculty/staff time for admissions work, program assessment and program support such as faculty advising. For faculty, this may come in the form of course buyout, summer salary, or unrestricted funds. Only minor additional burden will be placed on other campus resources. In particular, the students will not enroll in ESL courses since their TOEFL score will be sufficiently high (at least 92) that they will not be required to take ESL.

13.2 Budget Detail

Students in this program revenue program are not eligible for any TA, LSA, RA or PA support. Faculty instructors for MS-DS courses have their traditional teaching likely covered by LSAs supported through this budget. TAs and LSAs identified in this MS-DS budget will support graduate students in traditional programs. The following are the main budget items (see separate spreadsheet for details) for a 20-student cohort:

- staff coordination ($5K)
- faculty coordinator ($14K for 0.5 summer month + 33% academic buyout)
- 2.67 LSAs, 0.67 TAs, 1.67 Sr TAs ($101K)

This totals $120K expenses, to be offset by income. Income, based on ¾ of 80% of tuition, would be $356K for 100% non-residents, $248K for 50% non-residents, $194K for 25% non-residents, or $140K for 0% non-residents.

14. Overlap with Other Programs

There is no overlap with other Department or University programs. However, new courses inspired by this MS-DS degree will be useful for existing traditional degree programs.

15. Assessment and Program Review Plan

This data science program will involve cohorts of students in hands-on big data experiences, where they will 1) acquire basic data science knowledge (design, sampling, visualization, inference); 2) develop skills with computational tools (data management, distributed computing); and 3) grasp challenges of complex data (not just using tools).

Success will be seen in a workforce with enhanced leadership and communication skills concerning data science, able to 1) direct/lead/inform management about the big picture, future directions, key needs; 2) share with lay audience about technical issues, visualization of trends; and 3) discuss deep issues with other technical experts in order to solve complex problems. In summary, students will develop a balanced perspective on extracting and sharing useful information from data.

In addition, success will be measured by enhancements to our traditional program, as we evolve our curriculum towards more relevant and pragmatic topics. The Department of Statistics has a long tradition of balancing theory and practice. This program provides a unique opportunity to better connect our research excellence in data science with our instructional mission. Further, we expect that the new connections with industry will lead to novel collaborations that will open up new lines of research over time.

Selected learning goals. The key learning objectives are that graduates will be able to

- define key aspects of a problem and conduct appropriate data science, summarizing results with insightful visuals for a broad audience;
- effectively convey the results of data science to decision makers orally and in writing;
- express the complexities and context dependencies, and pros and cons of data analysis decisions rather than simply presenting and explaining one possible solution; and
- understand the scope of a big data problem, scaling down for initial grasp of patterns and methods of data science, and scaling up for production to complete a project.
Assessment of learning. Traditionally the department has had a Course Evaluation Committee. We will initiate an Assessment Committee with 3-4 members of the faculty/staff, and ask the committee to assess the Master’s program as part of its work, in collaboration with the Graduate Program Committee.

The following are some ways in which we will assess student learning and preparation for admission to Ph.D. programs:

• We will monitor the success of the students in the beginning required courses. Based on the performance on exams we will assess whether the students can meet the goals outlined above. Instructors will provide input on how the Master’s students perform in comparison with advanced undergraduates in those courses.
• We will compare the performance of the Master’s students with the performance of advanced undergraduates and Ph.D. students in the introductory graduate classes.
• We will track the placement of our graduates into industry or government and into Ph.D. programs.
• We will conduct exit surveys upon graduation.

Revenue expectations. The department will monitor the program to ensure that it continues to meet the expectations of revenue. The financial success of the program will at least initially depend on the success of the anticipated collaborative program with several international universities with which faculty have established relationships.

Program Review. We propose that the program be monitored carefully during the first five years. A formal program review will be written after five years. If the program is considered successful, it will then be reviewed within the ten-year cycle of regular program review.

16. Summary

The proposed Master of Science in Statistics option in Data Science has been designed to meet the needs of a growing population of workers seeking skills in data science. We expect many of these students will come from industry and government in the U.S., wanting to upgrade skills. The Department has presented a detailed description of the program requirements and curriculum, the academic and administrative support structures, the budget implications and the evaluation plan. It is our hope that the program will be approved on a timeline so that students can be admitted for the academic year 2015-2016. The addition of the Master’s degree will mean that the UW-Madison is able to offer a professional degree to non-traditional students in addition to its already successful degree programs at all levels, training undergraduate Stat Majors, Master’s students and Ph.D. students. We anticipate a stronger overall department with vertical integration among the student groups as a result of this diversification. There are also horizontal benefits for other students outside this MS-DS program who are in urgent need for pragmatic data science courses.