FOR APPROVAL	PUBLIC	OPEN SESSION

TO: Committee on Academic Policy and Programs

SPONSOR: Sioban Nelson, Vice-Provost, Academic Programs
CONTACT INFO: (416) 978-3742, vp.academicprograms@utoronto.ca

PRESENTER: See Sponsor
CONTACT INFO:

DATE: December 6, 2017 for January 11, 2018

AGENDA ITEM: 4

ITEM IDENTIFICATION:

Proposal for a new Specialist in Data Science within an existing degree, Faculty of Arts and Science.

JURISDICTIONAL INFORMATION:

The Committee on Academic Policy and Programs approves new undergraduate programs within an existing degree, as defined by the University of Toronto Quality Assurance Process. *(AP&P Terms of Reference, Section 4.4.b.i).*

GOVERNANCE PATH:

1.   Committee on Academic Policy and Programs [for approval] (January 11, 2018)

PREVIOUS ACTION TAKEN:

The proposal for the Specialist in Data Science received approval from the Faculty of Arts and Science on December 13, 2017.

HIGHLIGHTS:

The proposed Specialist in Data Science was developed through the collaboration of the Departments of Computer Science and Statistical Sciences, Faculty of Arts and Science. The Specialist partially fulfils the requirements of a Bachelor of Science degree. Admission will be indirect and limited, and students will complete either 13.0 or 13.5 full-course equivalents (FCEs), depending on course selection, to finish the program.
The proposed Specialist in Data Science has three fundamental aspects that are highly integrated. One aspect involves the acquisition of advanced expertise in statistical reasoning, methods, and inference, which are essential for any data analyst. A second aspect involves training in computer

science: the design and analysis of algorithms and data structures for handling large amounts of data, and best practices in software design. Students will also receive training in machine learning, which lies at the intersection of computer and statistical sciences. The third aspect is the application of computer science and statistics to produce analyses of complex, large-scale data sets, and the communication of the results of these analyses. Students will receive training in these three areas through newly developed integrative and capstone courses in the program. The successful student will combine their expertise in computer and statistical science to produce and communicate analyses of complex, large-scale data sets.

The curriculum consists of existing courses in the Computer Science and Statistical Sciences programs as well as three new integrative courses in Data Science, one to be taken in Years 2 through 4. The final Data Science course will include a capstone project. The program is structured so that choice and elective room increase as students progress through the program.

The proposed program responds to the demand for those with the capacity to curate and analyze data as well as to think critically about the uses and abuses of big data. There is a high demand in data-intensive research fields and commercial endeavours for students completing undergraduate degrees in computer or statistical sciences who can think algorithmically as well as statistically. These students are expected to understand how to utilize databases and other data warehouses, scrape data from Internet sources, and program solutions to complex problems in multiple languages. Graduates of this program should be in a position to obtain employment in industry and government, where there is extraordinary demand for data scientists. They may also pursue graduate studies in computer science, statistics, or related fields.

Consultation across the University took place with the Faculty of Applied Science and Engineering, University of Toronto Mississauga, and University of Toronto Scarborough.

The program was subject to external appraisal on September 22, 2017 by Professor Maneesh Agrawala, Department of Computer Science, Stanford University and Professor John Perkau, Department of Statistics, University of British Columbia. The external appraisers made a number of suggestions which resulted in changes to the program, as reflected in the Dean's response to the appraisal report.

FINANCIAL IMPLICATIONS:

The new financial obligations resulting from this program will be met at the Faculty level.

RECOMMENDATION:

Be it Resolved

THAT the proposed Specialist in Data Science, which will confer the existing degree BSc, as described in the proposal from the Faculty of Arts and Science dated November 9, 2017, be approved effective September 1, 2018.

DOCUMENTATION PROVIDED:

- *Proposal for the Creation of the Specialist in Data Science*

University of Toronto
New Undergraduate Program Proposal

(This template has been developed in line with the University of Toronto's Quality Assurance Process.)

This template should be used to bring forward all proposals for new undergraduate programs for governance approval under the University of Toronto's Quality Assurance Process. It is designed to ensure that all evaluation criteria established by the Quality Council are addressed in bringing forward a proposal for a new program.

Please note that all proposed new undergraduate programs are subject to external review.

Name of Proposed Program:	**Specialist in Data Science**
Degree conferred:	**Honours B.Sc.**
Department / Unit (if applicable) where the program will be housed:	**Departments of Computer Science and Statistical Sciences**
Faculty / Academic Division:	**Faculty of Arts and Science**
Dean's Office Contact:	**Professor Pamela Klassen, Vice-Dean, Undergraduate & International; Professor Mary Pugh, Acting Vice-Dean, Undergraduate & International**
Proponent:	**Professors Ravin Balakrishnan and Jamie Stafford**
Direct entry or selection of POSt at end of 1st year:	**Selection of POSt at end of 1st year**
Version Date:	**November 9, 2017**

New Undergraduate Program Proposal

Undergraduate Specialist in Data Science
Departments of Computer Science and Statistical Sciences
Faculty of Arts and Science

Table of Contents

1 Summary

A new specialist program in Data Science is being jointly proposed by the Departments of Computer Science and Statistical Sciences, which will consist of 13.5 full course equivalents (FCEs) and will partially fulfill the requirements for the conferral of the Bachelor of Science degree. The proposed program in Data Science has three fundamental aspects that are highly integrated. One aspect involves the acquisition of advanced expertise in statistical reasoning, methods, and inference, which are essential for any data analyst. A second aspect involves training in computer science: the design and analysis of algorithms and data structures for handling large amounts of data, and best practices in software design. Students will also receive training in machine learning, which lies at the intersection of computer and statistical sciences. The third aspect is the application of computer science and statistics to produce analyses of complex, large-scale datasets, and the communication of the results of these analyses. Students will receive training in these three areas through newly developed integrative and capstone courses in the program. The successful student will combine their expertise in computer and statistical science to produce and communicate analyses of complex large-scale datasets.

Specific skills that graduates of the program will acquire include: proficiency in statistical reasoning and computational thinking; data manipulation and exploration, visualization, and communication that are required for work as a data scientist; the ability to apply statistical methods to solve problems in the context of scientific research, business, and government; familiarity and experience with best practices in software development; and knowledge of current software infrastructure for handling large data sets. Graduates of the program will be able to demonstrate the ability to: apply machine learning algorithms to large-scale datasets that arise in scientific research, government, and business; create appropriate data visualizations for complex datasets; identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data, and communicating the results; present the results and limitations of data analysis projects at an appropriate technical level for the intended audience.

Admission to the program will be competitive and enrolment will be limited. We anticipate an enrolment of around 110 students by the 2023-2024 academic year. We expect the demand for this program to be greater than the number of available spots since the demand for trained data scientists and the enrolments in both computer and statistical sciences have been increasing during the past several years. The program will be limited to 30 in the 2018-2019 academic year.

There is a high demand in data intensive research fields and commercial endeavours for students completing undergraduate degrees in computer or statistical sciences who can think algorithmically as well as statistically. These students are expected to understand how to utilize databases and other data warehouses, scrape data from Internet sources, and program

solutions to complex problems in multiple languages (Hardin et al., 2015).[1] Indeed, as science itself becomes a body of data that we can analyze and study, there are staggeringly large opportunities for improving the accuracy and validity of science, through data science – the scientific study of data analysis (Donoho, 2015).[2]

A committee was formed in June 2015 to advise on the development of this program. Committee members include seven U of T faculty members from both Statistical Sciences and Computer Science, and data scientists working at U of T affiliated institutions and private industry. An advisory committee of nine faculty members from Statistical Sciences and Computer Science and data scientists working in industry was also established to help the committee in developing this specialist program.

2 Effective Date

Anticipated date students will start the program: **September 2018**

3 Program Rationale

Data Science is an emerging discipline, which we seek to acknowledge in the program name. The proposed program in Data Science has three fundamental aspects that are highly integrated. One aspect involves the acquisition of advanced expertise in statistical reasoning, methods, and inference essential for any data analyst. A second aspect involves training in computer science: the design and analysis of algorithms and data structures for handling large amounts of data, and best practices in software design. Students will receive training in machine learning, which lies at the intersection of computer and statistical sciences. The third aspect is the application of computer science and statistics to produce analyses of complex, large-scale datasets, and the communication of the results of these analyses; students will receive training in these areas through newly developed integrative and capstone courses in the program. The successful student will combine their expertise in computer and statistical science to produce and communicate analyses of complex large-scale datasets, for example, applying machine learning to forecast crop yields based on climatic parameters.

As data become ubiquitous and easier to acquire, particularly on a massive scale, the demand for expertise in both computer and statistical science has become acute; this has resulted in

---

[1] Hardin, J. et al. (2015). Data science in statistics curricula: Preparing students to "Think with Data". *The American Statistician*, 69(4), 343-353.

[2] Donoho, D. (2015). 50 years of Data Science. In *Princeton NJ, Tukey Centennial Workshop*.

dramatic enrolment increases in these disciplines across North America. This trend has been driven by the growth of computational power, networks and the pervasive use of digital technologies. In many areas of science and industry there is an urgent demand for expertise in managing, mining, analyzing and interpreting rich data. The proposed program in data science is designed to provide students with a skill set in computer and statistical sciences to meet the demand for data scientists. Most current undergraduate programs at peer institutions in North America are inadequate for training data scientists: they offer training in either computer science or statistical science, but not both and they offer limited or no training in analyzing complex large-scale datasets and communicating the results of the analysis.

The demand for training in the statistical sciences at the interface with computer science is not a fleeting trend. It has led to an entirely new type of professional – the Data Scientist. Society's increasing dependence on data, and the emergence of large-scale data in many fields of endeavour, has led to a demand for data scientists that will persist. A data scientist works on the management, visualization, curation and analysis of large and complex datasets. A well-educated data scientist should understand the demands and limits of Big Data, by developing interdisciplinary expertise as a collaborative scientist with a strong computational and statistical skill-set.

Recent pedagogical projects in the Departments of Computer Science and Statistical Sciences (with financial support from the Provost's office, the Faculty of Arts and Science, the Department of Statistical Sciences, and the Ministry of Advanced Education and Skills Development) have resulted in tremendous improvements in the learning experience in large-enrolment introductory computer science and statistics courses. This work includes developing highly successful MOOCs and MAESD-funded online learning modules in both Computer Science (for the courses "Introduction to Computer Programming," CSC108H1, and "Software Tools and Systems Programming," CSC209H1) and Statistical Sciences ("Introduction to the Practice of Statistics" STA220H1). Computer Science and Statistics have demonstrated the ability to manage large and innovative pedagogical projects, to see them successfully through to completion, and to evaluate their impact. Through this work, we have developed experience and expertise in online and blended learning, including inverted classrooms. These recent innovations will be used to deliver a student-focused curriculum that covers the foundational areas of data science.

The curriculum will have an optional internship or PEY (*Professional Experience Year*) component so that students have an opportunity to gain practical experience as a data scientist. Students will also have other experiential learning opportunities through the integrative courses.

The University of Toronto is not the only institution to recognize the demand for data science training, and new data science programs are being created at centres of higher learning across the country. However, there are few universities in Canada with a computer science department and a statistical science department large enough to offer the breadth and depth available to students at the University of Toronto: on the St. George Campus we offer three programs in Computer Science with 9 focuses for the Specialist program, eight programs in Statistics or Applied Statistics, along with co-op opportunities in both disciplines, and the Masters of Science in Applied Computing with a Data Science concentration. Compared to the collaborative

interdisciplinary approach to Data Science at the Faculty of Arts and Science at U of T, most programs at other universities are of limited quality or biased towards one discipline. For example, the Department of Computer Science at Dalhousie University offers a Bachelor's degree in computer science with a specialization in data science that gives only a cursory treatment of statistical inference, quantitative reasoning and statistical methodologies. Similarly, the Department of Mathematics and Statistics at Laval University offers a Bachelor's degree in statistics with a concentration in data science that gives cursory treatment to the fundamentals of computing. The University of Waterloo offers a Bachelor of Computer Science (Data Science) and Bachelor of Mathematics (Data Science). The required courses are a mix of computer science, math, and statistics courses with no specific courses in Data Science.

The situation in the United States is similar. In contrast, the proposed U of T program is a truly joint venture involving leading departments in Canada in both disciplines. The integrative courses provide focused in-depth training in the practice of data science beginning in the first year of the program and reinforced and deepened in each subsequent year. In addition, students will have an opportunity for an internship (through the PEY program or otherwise) that will give them further experiential learning opportunities. Our investigations into programs across North America reveal that there are no other undergraduate programs comparable to what we plan. The University of Toronto has a unique opportunity to play a leadership role and set a high standard for the expectations within such a program, based on our strengths in both computer and statistical sciences, as described above and elsewhere in this document (for example in section 12). These strengths include faculty with expertise in all areas of data science such as machine learning, data analysis, pedagogy in computer science and statistical science, and software development. In addition, several faculty members with expertise in data science are jointly appointed in the Departments of Computer Science and Statistical Sciences.

4 Need and Demand

The emergence of large-scale complex data in nearly every facet of academic and daily life has been accompanied by an increasing demand for expertise at the interface of computer and statistical science. There has been a massive increase in the amount of data available from a growing array of new technologies. New data sources, such as network data, image data, and streaming data, are part of a Big Data revolution that has swept both disciplines and daily life. This trend will only intensify and must be met with parallel developments in the training of undergraduate students in data science with the capacity to curate and analyze date, as well as to think critically about the uses and abuses of big data. The demand for such training is certainly reflected in the increasing enrolments of both the Computer Science and Statistical Sciences departments. We also see it in the number of placements available for graduate and undergraduate students who are studying in either department: the PEY office posted almost 50 positions related to Data Science last year (even though there are currently few potential candidates with those skills by the end of their third year of undergraduate studies) and the Director of the Professional Master's program in Computer Science reports that for 2016–17, the

number of internship positions in Data Science exceeded the number of candidates from the program. The proposed program is a direct response to this demand. Graduates from this program should be in a position to obtain employment in industry and government, where the demand for data scientists is extreme. The McKinsey Global Institute estimates that the United States alone will face a shortage of about 140,000–190,000 graduates with Data Science skills in 2018 (Manyika et al., 2011).[3] Data Science graduates may also pursue graduate studies in computer science, statistics or related fields.

There is a high demand in data intensive research fields and commercial endeavours for students completing undergraduate degrees in computer or statistical sciences who can think algorithmically as well as statistically. These students are expected to understand how to utilize databases and other data warehouses, scrape data from Internet sources, and program solutions to complex problems in multiple languages (Hardin et al., 2015). Indeed, as science itself becomes a body of data that we can analyze and study, there are staggeringly large opportunities for improving the accuracy and validity of science, through data science – the scientific study of data analysis. (Donoho, 2015)

The program is distinct from other undergraduate programs at U of T for several reasons.

1. There is currently no other undergraduate program at U of T that covers the activities of the emerging field of Data Science, including: data exploration and preparation; data representation and transformation; computing with data; data modelling; data visualization and presentation; and the theory of data science (Donoho, 2015) – the rigorous study of data analysis performed on complex large-scale datasets.

2. To become trained as a data scientist, it is not sufficient for students to enrol in the existing computer science and statistics major programs, because a double major will not give them the benefit of an integrated pathway of courses designed explicitly to teach data science theories and methods. Only the Data Science specialist will give them an integrated experiential learning component, which will cover specialized Data Science topics such as data preparation, computational consideration involved in the statistical analysis of large-scale data sets, and the theory of data science. These experiential learning outcomes will be achieved through the integrative courses of the proposed program.

Canadian Universities Offering Undergraduate Degrees in Data Science

-   As mentioned above (see section 3), the Department of Computer Science at Dalhousie University and the Department of Mathematics and Statistics at Laval University, both offer their own data science programs from the perspective of their single disciplines. The University of Waterloo offers two separate programs in Data Science: B.Math. (Data

---

[3]   Manyika, J. et al. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute report.* McKinsey & Company.

Science) and B.CS. (Data Science). Unlike these degrees, the U of T Data Science Specialist will offer a sequence of courses that teach the foundations of Data Science by integrating topics in computer science, statistics, and elective courses with experiential learning opportunities. Students in the U of T specialist in Data Scientist will have opportunities during their degree to implement and communicate the rationale behind every step that a professional Data Scientist must take to learn from data, to analyze it, and to communicate results to non-specialists in work environments.

Non-Canadian Universities Offering Undergraduate Degrees in Data Science

- The situation in the United States is similar to Canada, in that Data Science programs are very limited. As a truly joint venture involving the leading Computer Science and Statistics departments in Canada, the U of T Data Science specialist stands out for its integrative courses that provide in-depth training in data science. Our investigations into programs across North America reveal that UC Irvine's B.A. program in Data Science is closest to our own highly integrative model, through their inclusion of a final year capstone project course that teaches students how to apply statistical and computational principles to solve large-scale, real-world data analysis problems. There are some M.A. programs that take a similar approach to our proposal, but even many of these programs are limited by the same bifurcated disciplinary approach described above. Several leading universities are in the early stages of developing Data Science programs. Berkeley has developed a blueprint for a comprehensive Data Science program but does not currently offer an undergraduate degree in Data Science. Harvard recently announced a new track in Data Science, although no details have been published. Princeton University has a Certificate in Statistics and Machine Learning where students take 5 one-semester courses in statistics, machine learning, and a concentration area such as the life sciences or economics. In addition, these "students are … required to complete a thesis or at least one semester of independent work in their junior or senior year on a topic that makes substantial application or study of machine learning or statistics."[4] Warwick University offers a three-year undergraduate degree in Data Science that includes an experiential learning component in the last year of the program, although there does not seem to be a series of courses that teach the foundations of Data Science throughout the degree. Based on this review of Data Science curricular initiatives at other institutions, our proposed program leverages our strengths in computer and statistical sciences, as well as the excellent collaborative relationship between these two departments, to offer something truly unique.

Table 1: Undergraduate Enrolment Projections

Level of	2018-2019	2019-2020	2020-2021	2021-2022*	2022-2023	2023-2024

---

[4]   http://sml.princeton.edu/certificaterequirements

study						
1st year						
2nd year	30	35	37	42	42	42
3rd year	-	26	31	33	37	37
4th year	-	-	22	26	28	31
Total enrolment	30	61	90	101	107	110

*Steady State

The table above provides the undergraduate enrolment projections. It is anticipated that enrolment will consist of about 74% domestic students and 26% international students.

The students will be drawn from existing enrolment streams. This program is not planned to increase Faculty enrolment. Both Computer Science and Statistics are highly sought-after with strong demand from our existing students, not all of these students are able enter their program of choice. This proposed program expands the supply of high-demand program offerings for our students by increasing Computer Science and Statistics enrolment, from existing Arts and Science enrolments. The costs associated with the program are aligned with those of a Computer Science education, funded from student tuition and government grant.

5 Admission Requirements

Students will be admitted to the specialist program in Data Science at the end of their first year. Enrolment in the program will be limited, as is currently the case with Computer Science programs, with admission to the program based on the students' grades in three first-year courses.

The program aims to attract and admit students who will be better at statistics than most software engineers and better at software engineering most statisticians. To identify good potential software engineers we will use CSC148H1 ("Introduction to Computer Science"), and to identify good potential statisticians, we will use STA130H1 ("An Introduction to Statistical Reasoning and Data Science") and MAT137Y1/157Y1 ("Calculus"). Students must have at least 70% in each of these courses. An annual cut-off will be established using applicants' grades in these courses. MAT157Y1 grades will be adjusted to account for the increased difficulty. The exact cut-offs used will vary year-to-year depending on the program's capacity and the cohort of applicants.

6 Program Requirements

The proposed data science program will provide graduates with a deep understanding of statistical and machine learning methods, programming and software development, algorithms and data structure. They will also develop expertise in communicating and visualizing data, including to non-specialists, and skills in thinking critically about the possibilities and limits of big data.

Specific skills that graduates of the program will acquire include: proficiency in statistical reasoning, and computational thinking; data manipulation and exploration, visualization, and communication that are required for work as a data scientist; the ability to apply statistical methods to solve problems in the context of scientific research, business, and government; familiarity and experience with best practices in software development; and knowledge of current software infrastructure for handling large data sets. Graduates of the program will be able to demonstrate the ability to: apply machine learning algorithms to large-scale datasets that arise in scientific research, government, and business; create appropriate data visualizations for complex datasets; identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data, and communicating the results; present the results and limitations of a data analysis at an appropriate technical level for the intended audience.

The total number of courses that students are required to take is 13.0–13.5 FCEs. Students are required to take: 3.0 FCEs during their first year; 3.5–4.0 FCEs during their second year (3.5 FCEs if they take the recommended CSC240H1; 4.0 if they take CSC165H1 and CSC236H1 instead), including the first integrative course; 3.5 FCEs during their third year including the second integrative courses; and 3.0 FCEs during their fourth year including the capstone course.

The new integrative courses are described below:

JSC270H1 Data Science I (new course)

This course is intended for students in the Data Science Specialist program. Data exploration and preparation; data visualization and presentation; and computing with data will be introduced. Professional skills, such as oral and written communication, and ethical skills for data science will be introduced. Data science workflows will be integrated throughout the course. These topics will be explored through case studies and collaboration with researchers in other fields.

JSC370H1 Data Science II (new course)

This course is intended for students in the Data Science Specialist program. Data representation and transformation; predictive and generative models. Professional and teamwork skills, such as oral and written communication, and ethics for data science. Data science workflows will be integrated throughout the course. Topics will be explored through case studies and collaboration with researchers in other fields.

JSC470H1 Data Science III (new course)

This course is intended for students in the Data Science Specialist program. Through case studies and collaboration with researchers in other disciplines, students develop skills in collaborating as a data scientist. The focus is on developing software that produces data analyses.

The program is structured so that choice and elective room increases as students' progress through the program. During the first year students have a choice in the theoretical level of calculus (MAT137Y1 – Calculus or MAT157Y1 – Analysis I), and linear algebra (MAT223H1 – Linear Algebra I or MAT240H1 – Algebra I). During the second year students have a choice in the theoretical levels of computational theory (CSC165H1 – Mathematical expression and reasoning, CSC240H1 – Enriched Introduction to the theory of computation) and data structures course (CSC263H1 – Data Structures and Analysis or CSC265H1 – Enriched Data Structures and Analysis). During the fourth year students can choose one of STA414H1 – Statistical Methods for Data Mining and Machine Learning or CSC411H1 – Machine Learning and Data Mining; and students will choose three courses from the following list of computer science and statistics courses, including at least two 400-level courses: CSC421H1 – Introduction to Neural Networks and Machine Learning, STA303H1 – Methods of Data Analysis II, STA347H1 – Probability I, CSC401H1 – Natural Language Computing, CSC412H1 – Probabilistic Learning and Reasoning, any 400-level STA course.

 A full list of courses is outlined in appendix A.

7 Degree Level Expectations and Learning Outcomes

A student graduating from the program will have met general criteria for the University of Toronto Faculty of Arts and Science Honours Bachelor of Science degree-level expectations. These degree level expectations are outlined below.

1. Depth and Breadth of Knowledge

Breadth of Knowledge

Students will be introduced to a wide variety of approaches to the study of Data Science. They will take elective courses in the Faculty of Arts and Science to become familiar with other disciplinary fields and modes of scholarly inquiry. They will investigate how the study of computer science and statistics relates to other academic fields and areas of human creativity, and be asked to consider how Data Science interacts with other academic fields within the sciences, social sciences, and humanities. For example, they might take electives related to the ethics of privacy and security in the era of Big Data, or the politics of state surveillance and corporate manipulation through such tools as algorithms. Students will be expected to demonstrate an understanding of the current state of research in and knowledge about diverse

fields of Data Science. They will also develop the ability to apply learning from other related disciplines to Data Science.

Depth of Knowledge

Students will develop detailed knowledge of and experience in a specialized area of study within Data Science. In addition to completing the Data Science curriculum, students will be encouraged to choose a specialized program of study that will emphasize more highly developed knowledge and skills. Within their chosen program of study, students will develop the ability to critically evaluate new data science methods.

2. Knowledge of Methodologies

Students will learn methodologies related to the analysis of data, machine learning algorithms, data visualizations, data preparation, writing computer programs, and knowledge of current and software infrastructure for handling large data sets. They will also learn effective methods for communicating complex quantitative information in interactive and accessible ways.

3. Application of Knowledge

Students will be expected to apply the knowledge that they acquire in novel ways and to develop both practical and speculative applications of the Data Science methodologies they have learned. On the practical side, students will be expected to apply statistical and machine learning methods, such as classification and regression, to predict responses in various application domains. For example, in clinical medicine the outcome might be cancer recurrence, and in e-commerce the outcome might be the amount of a purchase. On the speculative side, students will apply existing theories in new ways or apply these theories to new applications. They will demonstrate the ability to make critical and imaginative use of scholarship and apply it to the area they are investigating. In short, students will develop sound and well-informed judgment about what can be learnt from data.

4. Communication Skills

Students will be expected to communicate about the process of extracting information from data in both oral and written form. This includes every step that s/he must take from getting acquainted with the data all the way to delivering results based in it.

5. Awareness of Limits of Knowledge

Students will develop the skills to identify limits of their own knowledge and ability through working on problems where there are many possible solutions. Students will be expected to not only appreciate uncertainty, but to quantify uncertainty in statistical terms. They will find that most knowledge obtained through data is in essence provisional, as it is always subject to ongoing and unceasing review, reinterpretation, and revision, especially in light of new data. Their awareness of the limits of knowledge will inform their critical engagement with prior and

existing ideas about information and knowledge. It will also help them to make well-informed critical judgments about how the theories and methods of data science may, or may not, apply to many fields of inquiry.

6. Autonomy and Professional Capacity

Students will be expected to be demonstrate many habits of mind such as persistence, thinking flexibly, metacognition, striving for accuracy, questioning and posing problems, applying past knowledge to new situations, working collaboratively, and remaining open to continuous learning. They will be expected to be independent learners within Data Science and other disciplines, which they will encounter as Data Scientists. In all that they undertake, they will exhibit behaviour that is consistent with the ideals of academic honesty, integrity, and social responsibility.

The learning outcomes for the program are outlined in a curriculum map provided as an appendix. (see appendix H)

8 Program Structure

First Year

The first year courses in computer science introduce students to computer programming, and the fundamentals of good practices in software development.

The first year courses in statistics introduce students to data analysis, statistical reasoning, teamwork, and oral and written communication skills.

Second Year

The second year courses in statistics introduce statistical reasoning through the mathematical foundations of probability and statistics that are necessary for future study.

The second year courses in computer science will introduce essential algorithms and data structures, the fundamentals of reasoning about algorithms and data structures for processing data efficiently; reinforce the students' understanding of good practices in software development; and will give students experience applying them. Students will be introduced to tools useful for processing large-scale datasets. The students' understanding of algorithms and data structures for processing data efficiently and the analysis of these algorithms and data structures will be expanded and reinforced.

The first integrative course in the program will introduce students to the construction and evaluation of data analysis and visualization. Students will learn how to identify and answer

questions that involve the application of statistical or machine learning methods to complex data, and approaches for communicating the results.

Third Year

The third year courses in statistics will reinforce the critical evaluation and application of statistical methods and study design to develop solutions to problems based on data. The theory and methods of modern and classical statistical methods will be reinforced.

During the third year the students will gain competency and mastery in the design and analysis of algorithms and data structures for processing and storing large data sets efficiently.

The second integrative course will introduce students to the fundamentals of supervised and unsupervised learning methods. Students will apply their knowledge of the fundamental algorithms and data structures needed for storing and processing large amounts of data to answer questions based on data collected in the context of scientific research, business, and government. The answers to these questions will be communicated in oral and written form using data summaries such as visualizations and tables.

Fourth Year

During the fourth year courses in statistics students will gain competency and mastery in their understanding of the theory and application statistical inference, machine learning algorithms, modern and classical statistical methods. The students will gain competency and mastery in several advanced topics in statistics and/or machine learning.

The capstone course will allow students to consolidate their knowledge and skills towards becoming a competent data scientist. Students will create a data product – a piece of software that extracts insights from a dataset to produce usable results – that will solve a **real problem.** This will include the application of students' statistical and computer science skills that they have been developing throughout the program. Students should be able to apply or modify algorithms, including machine learning algorithms and data structures, to write computer programs to create a high quality data product that could be used in answering questions based on real data. Students should be able to communicate the results and limitations of their approach at the appropriate technical levels for the intended audience.

Table 2: Sample Pathway Towards Program Completion

Level of study	Fall Semester	Winter Semester
1st year	MAT137Y1 MAT223H1 CSC108H1	MAT137Y1 STA130H1 CSC148H1
2nd year	MAT237Y1 STA257H1 CSC207H1	MAT237Y1 STA261H1 CSC240H1 JSC270H1
3rd year	STA302H1 CSC209H1 CSC263H1 CSC343H1	STA303H1 STA355H1 JSC370H1
4th year	CSC373H1 STA414H1/CSC411H1 (**)	JSC470H1 (**)

(**) Additional 1.5 FCEs from the following list, including at least 1.0 FCE at the 400 level: STA303H1/STA305H1, STA347H1, CSC401H1, CSC412H1, CSC421H1, any 400-level STA course.

The table above provides a sample pathway for students applying after one year. It represents the expected "standard" path through the program.

Students applying after two years can make up the courses that they missed during the summer preceding their application. JSC270H1 is a prerequisite for JSC370H1, and JSC370H1 is a prerequisite for JSC470H1, and all three courses will currently only be offered in the Winter term. Therefore, three years will be required to complete these courses.

For example, students who take MAT135H1, MAT136H1, CSC108H1, CSC148H1, CSC165H1 during their first year (intending to pursue studies in Computer Science), can take MAT223H1 and STA130H1 during their first summer here (and MAT137Y1 as an extra credit if they did not achieve 90% or above in MAT136H1), to be ready to enter Data Science in Year 2. Similarly, students who take MAT137Y1, MAT223H1, STA130H1 during their first year (intending to pursue studies in Statistical Sciences), can take CSC108H1 during the summer, CSC148H1 during the fall and CSC207H1, CSC240H1 during the winter of their second year, to be ready for JSC270H1.

9 Assessment of Learning

Students in the program acquire a skill set that encompasses a foundation in Data Science. This foundation is assessed through standard means found in the required core courses of the program including the three integrative courses. Examples include standard tests and assignments to evaluate students' learning. Oral and written communication will be evaluated through writing assignments and presentations.

Assessment of the collaborative and experiential learning component of the program begins in the first integrative course, Data Science I (JSC270H1). This course will focus on: Data Exploration and Preparation; Data Visualization and Presentation; and Computing with Data. The second integrative course, Data Science II (JSC370H1), will focus on Data Representation and Transformation and Data Modelling with an emphasis on predictive modelling. The final course, Data Science III (JSC470H1), would include a capstone project, covering all six divisions, including the theory of Data Science (Donoho, 2015). Computing with Data would be a key feature of all 3 courses, with increasing sophistication over the 3 courses. The three courses will cover all six data science divisions with increasing sophistication

Assessments in the integrative courses will mainly include written assignments, individual and group presentations, and collaboration with peers, teaching assistants, and external collaborators. Attendance, participation, and preparation for class meetings will be evaluated. These courses will scaffold a student's ability to use Data Science methods to learn from data for practical problems. JSC270H1, JSC370H1, and JSC470H1 will consist of problems where applying knowledge from other courses will occur and they will also involve problems where there are many possible solutions. There will also be many opportunities for students to demonstrate autonomy and professional capacity. Ethical practice of data analysis and software development will be taught in these courses. All three courses will include experiential learning components with increasing sophistication over the three years.

The program's learning objectives as outlined in the curriculum map will be continually assessed throughout the program. Available data on specific learning objectives directly related to DLE will be monitored. The program director will monitor course evaluations, course grades, and GPA. As another form of assessment, we will survey data scientists and industry professionals that are part of the integrative courses, interact with students, and see their work first hand. We will also receive qualitative feedback on the program on a continuous basis from these professionals. This will help keep methods and topics in these courses current.

10 Consultation

A committee was formed in June 2015 to advise on the development of this program. Committee members include seven U of T faculty members from both Statistical Sciences and Computer Science, and data scientists working at U of T affiliated institutions and private industry. An advisory committee of nine faculty members from Statistical Sciences and

Computer Science, and data scientists working in industry was also established to help the committee in developing this specialist program.

A survey of managers, executives and practitioners in the field of Data Science was conducted by a sub-committee to learn about current industrial needs of data scientists. The results of the survey indicate that graduates of the program will meet the needs of companies planning to hire data science professionals.

An earlier draft of this proposal was circulated to all chairs sitting on the Sciences Curriculum Committee, and received favourable feedback. The brief is also circulating to all units within Arts and Science that have a potential interest in Data Science. In addition to informing these units of the proposed program we are also inquiring about the extent to which the Data Science Specialist could support data intensive studies in their unit through specific courses. Our intention is to encourage students in the Data Science specialist program to consider taking data intensive courses, as well as courses offering critical perspectives on the social, political, and economic effects of big data, from departments across Arts and Science. We will also recommend that students consider pursuing an accompanying minor in an area complementary to the Data Science Specialist.

We have also shared this proposal with the Faculty of Applied Science and Engineering and the Undergraduate Vice-Deans at UTM and UTSC for their feedback.

11 Resources:

Faculty in the Departments of Statistical Sciences and Computer Science will teach most courses in this program.

A faculty member in computer science or statistics will take on the role of program director of undergraduate Data Science. The primary role of the program director will be to seek, facilitate, and coordinate internship or PEY opportunities, and advise students in the program. The program director will require administrative support for program admissions and external relations with researchers and data scientists that will be involved in the three integrative courses. Given the small size of the program (roughly 40 students in each year), each of the integrative courses can be offered in a single section that will be staffed directly from the existing (and currently planned) complement of faculty members in each department.

We are proposing three new courses, one in the second, third, and fourth years of the program. An ATLAS application for pilot funding for teaching and learning innovation will be submitted this year to obtain resources for developing experiential learning modules in these three new courses.

The three new courses will be team taught by existing faculty in Computer Science and Statistics, who are practicing data scientists. There are active searches for a joint tenure-stream assistant professor position (joint between Statistical Sciences and Computer Science) in Data

Visualization,and for two teaching-stream assistant professor positions in Statistical Sciences. All three are expected to engage in teaching some of the courses required by the new program.

11.1 Faculty

The Departments of Computer Science and Statistical Sciences have faculty with research expertise in statistical theory, application of statistical and machine learning methods, programming and software development, algorithms and data structures, and statistical communication. Both departments are currently recruiting new faculty members in machine learning, data visualization, and statistical computation and these faculty members will also contribute to the program.

The Departments of Computer Science and Statistical Sciences host one of the world's leading machine learning research groups. The Dept. of Statistical Sciences hosts leading researchers in theoretical statistics, applied statistics, and statistical computation. Members of the Dept. of Computer Science's Human-Computer Interaction team are leading researchers in interacting with and visualizing large amounts of data.

Both Departments host many researchers whose expertise is in analyzing data in specific domains, including medical and biological data, spatial data, and image data.

Teaching stream faculty in the Deptartments of Computer Science and Statistical Sciences have expertise in teaching undergraduate classes focused on both processing and analyzing large amounts of data and in teaching the communication skills that are required of Data Science professionals. Teaching stream faculty in both Departments engage in leading scholarship in pedagogy.

Both Departments have developed expertise in running academic programs with internship components, at the undergraduate and the graduate levels, and faculty will be able to support further developments of connections to industry in the context of the Data Science program.

The research-informed nature of the program will benefit from the faculty identified in the table below.

Table 3: Detailed Listing of Committed Faculty

Faculty name and rank	Home unit	Area(s) of Specialization
Ravin Balakrishnan, Professor and Chair	Computer Science	Human-computer interaction (HCI) with special focus on methods for interacting with and visualizing large amounts of data
Allan Borodin, University Professor	Computer Science	Mathematical foundations of Computer Science

Marzyeh Ghassemi, Assistant Professor	Computer Science and Medicine	Computational medicine: machine learning algorithms to predict and stratify human risks
Anna Goldenberg, Assistant Professor	Computer Science	Machine learning methods to decipher human disease heterogeneity
Roger Grosse, Assistant Professor	Computer Science	Machine learning
Matt Medland, Assistant Professor, Teaching Stream	Computer Science	Software Engineering
Francois Pitt, Associate Professor, Teaching Stream	Computer Science	Mathematical foundations of Computer Science
Khai Truong, Professor and Associate Chair Research	Computer Science	Ubiquitous computing (ubicomp) and human-computer interaction (HCI)
Raquel Urtasun, Associate Professor	Computer Science	Computer vision and machine learning, with a focus on machine perception for self-driving cars
Richard Zemel, Professor	Computer Science	Machine learning methods, with a specific focus on unsupervised learning, and probabilistic models of neural representations
Fanny Chevalier, Assistant Professor	Statistical Sciences and Computer Science	Data visualization and analytics
David Duvenaud, Assistant Professor	Statistical Sciences and Computer Science	Machine learning
Murat Erdogdu, Assistant Professor	Statistical Sciences and Computer Science	Machine learning, high-dimensional statistics
Radu Craiu, Professor	Statistical Sciences	MCMC, Statistical Computing, Genetics, Copula Modelling
Alison Gibbs, Associate Professor, Teaching Stream	Statistical Sciences	Statistical Education, Consulting

Keith Knight, Professor	Statistical Sciences	Robustness, Time Series
Nancy Reid, University Professor	Statistical Sciences	Likelihood, Experimental Design
Jeffrey Rosenthal, Professor	Statistical Sciences	MCMC, Stochastic Processes
Daniel Simpson, Assistant Professor	Statistical Sciences	Computational statistics, spatial statistics
Jamie Stafford, Professor	Statistical Sciences	Disease Mapping, Spatial Temporal Modelling
Lei Sun, Professor	Statistical Sciences	Statistical Genetics, False Discovery, Multiple Testing
Nathan Taback, Assistant Professor, Teaching Stream	Statistical Sciences	Statistical Education, Statistical Consulting, and Biostatistics.
Fang Yao, Professor	Statistical Sciences	Functional Data Analysis, Nonparametric Regression
Zhou Zhou, Associate Professor	Statistical Sciences	Functional Data Analysis, Time Series

11.2 TA Support

The integrative courses will be project-based and integrate ideas and techniques from computer science and statistics. These courses will have notable experiential learning components where students will: (1) gain experience working on data science projects that involve using data from industry, science, or the humanities to help answer salient questions; (2) interact with data scientists, researchers, or other professionals from academia or industry; (3) learn how to translate data science skills across domains and think critically about data and models of data; (4) develop strong oral and written communication skills and the ability to work in multidisciplinary teams. TA support that is higher than the usual level provided by the Faculty of Arts and Science for Computer Science and Statistics courses is required for each course.

11.3 Learning Resources

Please see the following Appendices.

Appendix C: Library statement confirming the adequacy of library holdings and support for student learning (attached)

Appendix D: statement concerning student support services (attached)

11.4 Space/Infrastructure

There is no space or infrastructure needed beyond that entailed by courses, all but three of which already exist. The three new courses will be able to be mounted within existing space allocated through the Faculty of Arts and Science.

12 Quality and other indicators

The Departments of Computer Science and Statistical Sciences currently have many faculty members specializing in research and pedagogy in all aspects of Computer and Statistical sciences that comprise Data Science. Both departments are experiencing significant faculty growth in machine learning, data visualization, and statistical computation that will continue to support the program. Statistical Sciences is currently reviewing its undergraduate curriculum and developing a detailed curriculum map to ensure that its program learning outcomes are well communicated to both faculty and students.

The Departments of Computer Science and Statistical Sciences host one of the world's leading machine learning research groups. For example, work done by Prof. Geoffrey Hinton and the University of Toronto's machine learning group is widely recognized as cutting edge research in the area of Deep Learning. Members of the group regularly teach introductory and advanced courses that will be part of the curriculum of the Data Science Specialist program. Research faculty in the machine learning group regularly renew and develop the undergraduate machine learning curriculum, and develop high-quality teaching materials used inside and outside the University. For example, Prof. Geoffrey Hinton's course on neural networks, originally developed for University of Toronto, is also available and widely-viewed on the Coursera MOOC platform.

The Department of Computer Science hosts a leading Computer Science Theory group. Members of the group regularly teach and participate in the design of courses on efficiently processing large amounts of data.

The Department of Statistical Sciences hosts leading researchers in all statistical aspects of Data Science. For example, in recent years, Prof. Jeffrey Rosenthal and Prof. Nancy Reid were awarded the Statistical Society of Canada (SSC) Gold Medals for outstanding contributions to Statistics. Prof Nancy Reid has made fundamental contributions to statistical inference. Prof. Jeffrey Rosenthal made fundamental contributions to statistical computation through his work

in Markov chain theory. Faculty in the Department of Statistical Sciences teach and develop courses that are part of the Data Science Specialist curriculum. For example, Prof. Jeffrey Rosenthal recently developed and taught the inaugural offering STA130H1 (An Introduction to Statistical Reasoning and Data Science), a required course in the Data Science Specialist curriculum. Prof. Nathan Taback has developed a modern experimental design course (STA305H1: The Design of Scientific Studies) that includes topics such as causal inference, another required course in the program. Prof. Lei Sun has developed a course in statistical genetics, STA480H1, that can be taken as part of the program for students who are interested in computational biology and data science.

Faculty in the Depts. of Computer Science and Statistical Sciences have expertise in teaching undergraduate classes on both processing and analyzing large amounts of data and in teaching the communication skills that are required of Data Science professionals.

Teaching stream faculty in both departments engage in scholarship related to the pedagogy of various aspects of data science and in course development. For example, Professors Nathan Taback and Alison Gibbs have both taught courses that involve teaching communication of statistical ideas to non-statisticians. Teaching stream faculty are involved in the development of all aspects of the undergraduate curriculum in cooperation and consultation with research stream faculty and develop teaching materials used inside and outside the University. For example, materials developed by Michael Guerzhoy for a University of Toronto course on machine learning and neural networks are also used by a Udacity MOOC.

The excellence in teaching, pedagogical scholarship and curricular innovation of teaching stream faculty in both Departments in has been recognized through external and internal awards. For example, Profs. Alison Gibbs (DoSS), Paul Gries (DCS), Diane Horton (DCS), and Karen Reid (DCS) have all received the prestigious University of Toronto President's Teaching Award in recent years. (See http://www.provost.utoronto.ca/Awards/presidentaward.htm).

Both Departments have developed expertise in running academic programs with internship components, both at the undergraduate and the graduate levels, alone and in cooperation with other units of the University. In 2015-2016, 178 undergraduate Computer Science students obtained 12-16 month internships through the Professional Experience Year (PEY) program. All the students in the Master of Science in Applied Computing (MScAC) program, including in the Data Science stream of the program, which is run by both Departments, are required to complete an 8-month internship. The MScAC program has excellent placement record, with more positions in data science than there are students. The expertise and industry contacts of the faculty and staff running the MScAC program can be used to benefit Data Science Specialist students.

13 Governance Process:

	Levels of Approval Required
Decanal and Provostial Sign-Off	March 2017
	Unit Level approval
	Faculty/Divisional Governance December 13, 2017
Submission to Provost's Office	
	AP&P – January 11, 2018
	Academic Board (if a new degree) NA
	Executive Committee of Governing Council (if a new degree) NA
Program may begin advertising as long as any material includes the clear statement that "No offer of admissions will be made to the program pending final approval by the Quality Council and the Ministry of Colleges Training and University (where the latter is required)."	
	Ontario Quality Council February 2018
	Submitted to MAESD (in case of a new degree) NA

Appendix A: Courses

CSC108H1 **Introduction to Computer Programming [36L]**

Programming in a language such as Python. Elementary data types, lists, maps. Program structure: control flow, functions, classes, objects, methods. Algorithms and problem solving. Searching, sorting, and complexity. Unit testing. No prior programming experience required.

CSC148H1 **Introduction to Computer Science [36L/24P]**

Abstract data types and data structures for implementing them. Linked data structures. Encapsulation and information-hiding. Object-oriented programming. Specifications. Analyzing the efficiency of programs. Recursion. This course assumes programming experience as provided by CSC108H1. Students who already have this background may consult the Computer Science Undergraduate Office for advice about skipping CSC108H1. Practical (P) sections consist of supervised work in the computing laboratory. These sections are offered when facilities are available, and attendance is required.

CSC165H1 **Mathematical Expression and Reasoning for Computer Science [36L/24T]**

Introduction to abstraction and rigour. Informal introduction to logical notation and reasoning. Understanding, using and developing precise expressions of mathematical ideas, including definitions and theorems. Structuring proofs to improve presentation and comprehension. General problem-solving techniques. Running time analysis of iterative programs. Formal definition of Big-Oh. Diagonalization, the Halting Problem, and some reductions. Unified approaches to programming and theoretical problems.

CSC207H1 **Software Design [24L/12T]**

An introduction to software design and development concepts, methods, and tools using a statically-typed object-oriented programming language such as Java. Topics from: version control, unit testing, refactoring, object-oriented design and development, design patterns, advanced IDE usage, regular expressions, and reflection. Representation of floating-point numbers and introduction to numerical computation.

CSC209H1 **Software Tools and Systems Programming [24L/12T]**

Software techniques in a Unix-style environment, using scripting languages and a machine-oriented programming language (typically C). What goes on in the operating system when programs are executed. Core topics: creating and using software tools, pipes and filters, file processing, shell programming, processes, system calls, signals, basic network programming.

CSC236H1 **Introduction to the Theory of Computation [24L/12T]**

The application of logic and proof techniques to Computer Science. Mathematical induction; correctness proofs for iterative and recursive algorithms; recurrence equations and their solutions; introduction to automata and formal languages. This course assumes university-level

experience with proof techniques and algorithmic complexity as provided by CSC165H1. Very strong students who already have this experience (e.g., successful completion of MAT157Y1) may consult the undergraduate office about proceeding directly into CSC236H1.

CSC240H1 Enriched Introduction to the Theory of Computation [24L/12T]

The rigorous application of logic and proof techniques to Computer Science. Propositional and predicate logic; mathematical induction and other basic proof techniques; correctness proofs for iterative and recursive algorithms; recurrence equations and their solutions (including the Master Theorem); introduction to automata and formal languages. This course covers the same topics as CSC236H1, together with selected material from CSC165H1, but at a faster pace, in greater depth and with more rigour, and with more challenging assignments. Greater emphasis will be placed on proofs and theoretical analysis. Certain topics briefly mentioned in CSC165H1 or CSC236H1 may be covered in more detail in this course, and some additional topics may also be covered.

CSC263H1 Data Structures and Analysis [24L/12T]

Algorithm analysis: worst-case, average-case, and amortized complexity. Expected worst-case complexity, randomized quicksort and selection. Standard abstract data types, such as graphs, dictionaries, priority queues, and disjoint sets. A variety of data structures for implementing these abstract data types, such as balanced search trees, hashing, heaps, and disjoint forests. Design and comparison of data structures. Introduction to lower bounds.

CSC265H1 Enriched Data Structures and Analysis [24L/12T]

This course covers the same topics as CSC263H1, but at a faster pace, in greater depth and with more rigour, and with more challenging assignments. Greater emphasis will be placed on proofs, theoretical analysis, and creative problem-solving. Certain topics briefly mentioned in CSC263H1 may be covered in more detail in this course, and some additional topics may also be covered. Students without the exact course prerequisites but with a strong mathematical background are encouraged to consult the Department about the possibility of taking this course.

CSC343H1 Introduction to Databases [24L/12T]

Introduction to database management systems. The relational data model. Relational algebra. Querying and updating databases: the query language SQL. Application programming with SQL. Integrity constraints, normal forms, and database design. Elements of database system technology: query processing, transaction management.

CSC373H1 Algorithm Design, Analysis & Complexity [36L/12T]

Standard algorithm design techniques: divide-and-conquer, greedy strategies, dynamic programming, linear programming, randomization, network flows, approximation algorithms. Brief introduction to NP-completeness: polynomial time reductions, examples of various NP-complete problems, self-reducibility. Additional topics may include approximation

and randomized algorithms. Students will be expected to show good design principles and adequate skills at reasoning about the correctness and complexity of algorithms.

CSC401H1 Natural Language Computing [24L/12T]

Introduction to techniques involving natural language and speech in applications such as information retrieval, extraction, and filtering; intelligent Web searching; spelling and grammar checking; speech recognition and synthesis; and multi-lingual systems including machine translation. N-grams, POS-tagging, semantic distance metrics, indexing, on-line lexicons and thesauri, markup languages, collections of on-line documents, corpus analysis. PERL and other software.

CSC411H1 Machine Learning and Data Mining [24L/12T]

An introduction to methods for automated learning of relationships on the basis of empirical data. Classification and regression using nearest neighbour methods, decision trees, linear models, and neural networks. Clustering algorithms. Problems of overfitting and of assessing accuracy. Problems with handling large databases. Emphasis on statistical approaches.

CSC412H1 Probabilistic Learning and Reasoning [24L/12T]

An introduction to probability as a means of representing and reasoning with uncertain knowledge. Qualitative and quantitative specification of probability distributions using probabilistic graphical models. Algorithms for inference and probabilistic reasoning with graphical models. Statistical approaches and algorithms for learning probability models from empirical data. Applications of these models in artificial intelligence and machine learning.

CSC421H1 Neural Networks and Deep Learning [24L/12P] – (NEW)

An introduction to neural networks and deep learning. Backpropagation and automatic differentiation. Architectures: convolutional networks and recurrent neural networks. Methods for improving optimization and generalization. Neural networks for unsupervised and reinforcement learning.

JSC270H1 Data Science I [24L/12T] – (NEW)

This course is intended for students in the Data Science Specialist program. Data exploration and preparation; data visualization and presentation; and computing with data will be introduced. Professional skills, such as oral and written communication, and ethical skills for data science will be introduced. Data science workflows will be integrated throughout the course. These topics will be explored through case studies and collaboration with researchers in other fields.

JSC370H1 Data Science II [24L/12T] – (NEW)

This course is intended for students in the Data Science Specialist program. Data representation and transformation; predictive and generative models. Professional and teamwork skills, such as oral and written communication, and ethics for data science. Data science workflows will be

integrated throughout the course. Topics will be explored through case studies and collaboration with researchers in other fields.

JSC470H1 Data Science III [24L/12T] – (NEW)

This course is intended for students in the Data Science Specialist program. Through case studies and collaboration with researchers in other disciplines, students develop skills in collaborating as a data scientist. The focus is on developing software that produces data analyses.

MAT137Y1 Calculus [72L/24T]

A conceptual approach for students with a serious interest in mathematics. Attention is given to computational aspects as well as theoretical foundations and problem solving techniques. Review of Trigonometry. Limits and continuity, mean value theorem, inverse function theorem, differentiation, integration, fundamental theorem of calculus, elementary transcendental functions, Taylor's theorem, sequence and series, power series. Applications.

MAT157Y1 Analysis I [72L/48T]

A theoretical course in calculus; emphasizing proofs and techniques, as well as geometric and physical understanding. Trigonometric identities. Limits and continuity; least upper bounds, intermediate and extreme value theorems. Derivatives, mean value and inverse function theorems. Integrals; fundamental theorem; elementary transcendental functions. Techniques of integration. Taylor's theorem; sequences and series; uniform convergence and power series.

MAT223H1 Linear Algebra I [36L/12T]

Systems of linear equations, matrix algebra, real vector spaces, subspaces, span, linear dependence and independence, bases, rank, inner products, orthogonality, orthogonal complements, Gram-Schmidt, linear transformations, determinants, Cramer's rule, eigenvalues, eigenvectors, eigenspaces, diagonalization.

MAT240H1 Algebra I [36L/24T]

A theoretical approach to: vector spaces over arbitrary fields, including C and Z_p. Subspaces, bases and dimension. Linear transformations, matrices, change of basis, similarity, determinants. Polynomials over a field (including unique factorization, resultants). Eigenvalues, eigenvectors, characteristic polynomial, diagonalization. Minimal polynomial, Cayley-Hamilton theorem.

MAT237Y1 Multivariable Calculus [72L]

Sequences and series. Uniform convergence. Convergence of integrals. Elements of topology in R^2 and R^3. Differential and integral calculus of vector valued functions of a vector variable, with emphasis on vectors in two and three dimensional euclidean space. Extremal problems, Lagrange multipliers, line and surface integrals, vector analysis, Stokes' theorem, Fourier series, calculus of variations.

MAT257Y1 Analysis II [72L/48T]

Topology of R^n; compactness, functions and continuity, extreme value theorem. Derivatives; inverse and implicit function theorems, maxima and minima, Lagrange multipliers. Integration; Fubini's theorem, partitions of unity, change of variables. Differential forms. Manifolds in R^n; integration on manifolds; Stokes' theorem for differential forms and classical versions.

STA130H1 An Introduction to Statistical Reasoning and Data Science [24L/24P]

This course, intended for students considering a program in Statistical Sciences, discusses the crucial role played by statistical reasoning in solving challenging problems from natural science, social science, technology, health care, and public policy, using a combination of logical thinking, mathematics, computer simulation, and oral and written discussion and analysis.

STA257H1 Probability and Statistics I [36L/12T]

A mathematically rigorous introduction to probability, with applications chosen to introduce concepts of statistical inference. Probability and expectation, discrete and continuous random variables and vectors, distribution and density functions, the law of large numbers. The binomial, geometric, Poisson, and normal distributions. The Central Limit Theorem. (Note: STA257H1 does not count as a distribution requirement course).

STA261H1 Probability and Statistics II [36L/12T]

A rigourous introduction to the theory of statistical inference and to statistical practice. Statistical models, parameters, and samples. Estimators for parameters, sampling distributions for estimators, and the properties of consistency, bias, and variance. The likelihood function and the maximum likelihood estimator. Hypothesis tests and confidence regions. Examples illustrating statistical theory and its limitations. Introduction to the use of a computer environment for statistical analysis. (Note: STA261H1 does not count as a distribution requirement course).

STA302H1 Methods of Data Analysis I [36L]

Introduction to data analysis with a focus on regression. Initial Examination of data. Correlation. Simple and multiple regression models using least squares. Inference for regression parameters, confidence and prediction intervals. Diagnostics and remedial measures. Interactions and dummy variables. Variable selection. Least squares estimation and inference for non-linear regression.

STA303H1 Methods of Data Analysis II [36L]

Analysis of variance for one-and two-way layouts, logistic regression, loglinear models, longitudinal data, introduction to time series.

STA305H1 Design of Scientific Studies [36L]

Experiments vs observational studies, experimental units. Designs with one source of variation. Complete randomized designs and randomized block designs. Factorial designs. Inferences for contrasts and means. Model assumptions. Crossed and nested treatment factors, random effects models. Analysis of variance and covariance. Sample size calculations.

STA347H1 Probability [36L]

An overview of probability from a non-measure theoretic point of view. Random variables/vectors; independence, conditional expectation/probability and consequences. Various types of convergence leading to proofs of the major theorems in basic probability. An introduction to simple stochastic processes such as Poisson and branching processes.

STA355H1 Theory of Statistical Practice [24L/12P]

STA355H1 provides a unifying structure for the methods taught in other courses, and will enable students to read methodological research articles or articles with a large methodological component. Topics covered include statistical models and distributions; fundamentals of inference: estimation, hypothesis testing, and significance levels; likelihood functions and likelihood-based inference; prior distributions and Bayesian inference.

STA414H1 Statistical Methods for Data Mining and Machine Learning [36L]

Statistical aspects of supervised learning: regression with spline bases, regularization methods, parametric and nonparametric classification methods, nearest neighbours, cross-validation and model selection, generalized additive models, trees, model averaging, clustering and nearest neighbour methods for unsupervised learning.

Appendix B: Calendar Copy

Program Title: Specialist in Data Science

Level of Instruction: Undergraduate

Division: Arts & Science

Unit(s): Department of Computer Science, Department of Statistical Sciences

Description

The field of Data Science is a combination of statistics and computer science methodologies that enable 'learning from data'. A data scientist extracts information from data, and is involved with every step that must be taken to achieve this goal, from getting acquainted with the data to communicating the results in non-technical language. The Data Science Specialist program prepares students for work in the Data Science industry or government and for graduate studies in Data Science, Computer Science, or Statistics. Students in the program will benefit from a range of advanced courses in Computer Science and Statistics offered by the University of Toronto, as well as from a sequence of three integrative courses designed especially for the program.

The Data Science Specialist program comprises three fundamental and highly-integrated aspects. First, students will acquire expertise in statistical reasoning, methods, and inference essential for any data analyst. Seconds, students will receive in-depth training in computer science: the design and analysis of algorithms and data structures for handling large amounts of data, and best practices in software design. Students will receive training in machine learning, which lies at the intersection of computer and statistical sciences. The third aspect is the application of computer science and statistics to produce analyses of complex, large-scale datasets, and the communication of the results of these analyses; students will receive training in these areas by taking integrative courses that are designed specifically for the Data Science Specialist program. The courses involve experiential learning: students will be working with real large-scale datasets from the domain of business, government, and/or science. The successful student will combine their expertise in computer and statistical science to produce and communicate analyses of complex large-scale datasets.

Skills that graduates of the program will acquire include proficiency in statistical reasoning and computational thinking; data manipulation and exploration, visualization, and communication that are required for work as a data scientist; the ability to apply statistical methods to solve problems in the context of scientific research, business, and government; familiarity and experience with best practices in software development; and knowledge of current software infrastructure for handling large data sets. Graduates of the program will be able to demonstrate the ability to apply machine learning algorithms to large-scale datasets that arise in scientific research, government, and business; create appropriate data visualizations for complex datasets; identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data, and communicating the results; present the

results and limitations of a data analysis at an appropriate technical level for the intended audience.

Enrolment Requirements

This is a limited enrolment program (Type 2L) that can only accommodate a certain number of students. Eligibility is based on the following criteria:

A. Completion of at least 4.0 FCEs including CSC148H1 (with a minimum grade of 70%) and MAT137Y1/MAT157Y1 (with a minimum grade of 70%) and STA130H1 (with a minimum grade of 70%), AND
B. An average of the grades in CSC148H1 and MAT137Y1/MAT157Y1 and STA130H1 that meets the program's annual cutoff. MAT157Y1 grades will be adjusted to account for the course's greater difficulty. Note that the cutoff changes from year to year, depending on the current capacity of the program and the pool of applicants.

Note that students admitted to the program after second or third year will be required to pay retroactive program fees.

Completion Requirements

(13.0–13.5 Full Course Equivalents [FCEs], including at least 1.5 FCEs at the 400-level)

First year (3.0 FCEs)

MAT137Y1/MAT157Y1; MAT223H1/MAT240H1 (MAT240H1 is recommended); STA130H1; CSC108H1; CSC148H1;

Note: Students with a strong background in an object-oriented language such as Python, Java or C++ may omit CSC108H1 and proceed directly with CSC148H1. There is no need to replace the missing half-credit for program completion; however, please base your course choice on what you are ready to take, not on "saving" a half-credit. Consult with the Computer Science Undergraduate Office for advice on choosing between CSC108H1 and CSC148H1.

Second year (3.5–4.0 FCEs)

MAT237Y1/MAT257Y1; STA257H1; STA261H1; CSC207H1; (CSC165H1,CSC236H1)/CSC240H1 (CSC240H1 is recommended); JSC270H1 (Data Science I)

Note: CSC240H1 is an accelerated and enriched version of CSC165H1 plus CSC236H1, intended for students with a strong mathematical background, or who develop an interest after taking CSC165H1. If you take CSC240H1 without CSC165H1, there is no need to replace the missing half-credit for program completion; however, please base your course choice on what you are ready to take, not on "saving" a half-credit. Consult the Computer Science Undergraduate Office for advice on choosing between CSC165H1 and CSC240H1.

Third year (3.5 FCEs)

STA302H1; one of STA303H1 or STA305H1; STA355H1; CSC209H1; CSC263H1/CSC265H1 (CSC265H1 is recommended); CSC343H1; JSC370H1 (Methods of Data Science II)

Fourth year (3.0 FCEs)

1. CSC373H1;
2. one of STA414H1, CSC411H1;
3. JSC470H1 (Data Science III);
4. 1.5 FCEs from the following list, including at least 1.0 FCE at the 400 level: STA303H1/STA305H1 (whichever one was not taken in third year), STA347H1, CSC401H1, CSC412H1, CSC421H1, any 400-level STA course

Students will be advised to develop domain expertise in at least one area where Data Science is applicable, by taking a sequence of courses in that area throughout their program. Examples of such areas will be provided to students by program advisors and will form the basis for a later proposal for program Focuses (to be approved through internal Arts & Science governance procedures).

Appendix C: Library Report

**University of Toronto Libraries Report for the
Data Science Specialist Program
Department of Computer Science and Department of Statistical Sciences, 2016**

Context: The University of Toronto Library (UTL) system is the largest academic library in Canada and is currently ranked third among academic research libraries in North America, behind Harvard and Yale.[5] The research and special collections, together with the campus and college libraries comprise over 12 million print volumes, 5.6 million microform volumes, more than 17,000 journal subscriptions, in addition to a rich collection of manuscripts, films, and cartographic materials. The system provides access to more than 1.9 million electronic books, journals, and primary source materials and increasingly supports access via personal handheld devices.[6] There are numerous collection strengths in a wide range of disciplines reflecting the breadth of research and instructional programs at the University. The University of Toronto Library system has an annual acquisition budget of $31 million. The strong collections, facilities and staff expertise attract unique donations of books and manuscripts from around the world, which in turn draw scholars for research and graduate work.

Major North American Research Libraries[7]					
	2009-10	**2010-11**	**2011-12**	**2012-2013**	**2013-2014**
ARL RANK	**UNIVERSITY**	**UNIVERSITY**	**UNIVERSITY**	**UNIVERSITY**	**UNIVERSITY**
1	Harvard	Harvard	Harvard	Harvard	Harvard
2	Yale	Yale	Yale	Yale	Yale
3	**Toronto (3rd)**	**Toronto (3rd)**	**Toronto (3rd)**	**Toronto (3rd)**	**Toronto (3rd)**
4	Columbia	Michigan	Columbia	Columbia	Columbia
5	Michigan	Columbia	Michigan	Michigan	Michigan

Top 5 Canadian Universities in the ARL Ranking of Major North American Research Libraries				
2009-10	**2010-11**	**2011-12**	**2012-2013**	**2013-2014**
RANK/ UNIVERSITY	**RANK/ UNIVERSITY**	**RANK/UNIVERSITY**	**RANK/UNIVERSITY**	**RANK/UNIVERSITY**
3/Toronto	**3/Toronto**	**3/Toronto**	**3/Toronto**	**3/Toronto**
11/Alberta	11/Alberta	10/British Columbia	18/Alberta	22/British Columbia
24/British Columbia	16/British Columbia	15/Alberta	24/British Columbia	26/Alberta
31/Montreal	32/Montreal	18/McGill	30/McGill	35/McGill
37/McGill	38/McGill	32/Montreal	35/Montreal	36/Montreal

---

[5] Chronicle of Higher Education, "Library Investment Index at University Research Libraries, 2013 – 2014." In the Almanac of Higher Education, 2015. http://chronicle.com/article/Spending-by-University/232279

[6] Figures as of 2014 taken from UTL's 2015 Annual Report and 2013-2014 annual statistics. https://onesearch.library.utoronto.ca/sites/default/files/annual_reports/annualreport-2015.pdf and https://onesearch.library.utoronto.ca/annual-statistics/2013-2014

[7] Association of Research Libraries Statistics, 2013-14 http://www.arlstatistics.org/analytics

Space and Access Services: The Library system provides a variety of individual and group study spaces for both undergraduates and graduates in the 10 central and 23 divisional libraries on the St. George, Mississauga, Scarborough and Downsview campuses. Study space and computer facilities are available twenty-four hours, five days per week at one location, Robarts Library. Web-based services and electronic materials are accessible at all times from campus or remote locations, through the U of T based Scholars Portal and other leading edge digital services. The Map & Data Library offers additional personal computing facilities for accessing and analyzing data.

Instruction & Research Support: Libraries play an important role in the linking of teaching and research in the University. To this end, information literacy instruction shall be offered to assist in meeting the Data Science Specialist Program's degree-level expectations in the ability to gather, evaluate and interpret information. These services are aligned with the Association of College and Research Libraries (ACRL) Framework for Information Literacy for Higher Education[8].

Program Specific Instruction: Instruction can occur at a variety of levels for Data Science Specialist Program students and would be provided by faculty liaison librarians for Computer Science, Statistics and Data/GIS. Librarians in the Engineering & Computer Science Library, the Mathematical Sciences Library, and the Map & Data Library can facilitate formal instruction integrated into the class schedule and hands-on tutorials related to course assignments. The Library, through its liaison librarians, customizes feeds of library resources which appear prominently in Portal/Blackboard course pages. General subject area research guides can also be developed for students in the Data Science Specialist Program.

Librarians, data specialists, and GIS specialists in the Map & Data Library can offer web-based tutorials and in-person instruction on accessing and analysing data. In 2016-17, the Library intends to hire a Data Visualization Librarian to provide additional support in this area.

Collections: Many college and campus libraries collect materials in support of the Data Sciences Specialist Program; the main collections of these materials are housed in the Engineering & Computer Science Library and the Mathematical Sciences Library. Collections are purchased in all formats to meet the variety of preferences and styles of our current students and faculty. The University of Toronto Library is committed to collecting both print and electronic materials, including datasets, in support of the Data Science Specialist Program at the University of Toronto.

Journals: The Library subscribes to all of the top 25 journals listed in Journal Citation Reports (JCR) in the subject area of Statistics[9]. Of these titles, 25 are available electronically to staff and students of the University. 22 of the top 25 journals in Computer Science theory and methods are available electronically.

Monographs: The University of Toronto Library maintains comprehensive book approval plans with 53 book dealers and vendors worldwide. These plans ensure that the Library receives academic

---

[8]   Association of College & Research Libraries. Framework for Information Literacy for Higher Education. ACRL, 2016. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/infolit/Framework_ILHE.pdf

[9]   2015 Journal Citation Reports® (Thomson Reuters, 2015)

monographs from publishers all over the world in an efficient manner. In support of the Data Sciences Specialist Program, the Library specifically receives books through plans with Springer, Elsevier, Wiley, ACM, Cambridge University Press, Taylor & Francis, and Oxford University Press. In addition to these plans, individual librarian selectors select unique and interesting scholarly material overlooked by standard approval plans. These selections include individual e-books and e-book packages, contributions to the collections of the Thomas Fisher Rare Book Library, and special requests from faculty. In this way, the Library continues to acquire more than 120,000 book titles per year.

Preservation, Digitization, and Open Access: The University of Toronto Library supports open access to scholarly communication through its institutional research repository (known as T-Space), its open journal services, and subscriptions to open access publications. In addition to acquiring materials in support of programs in the Department of Computer Science and the Department of Statistical Sciences, the Library, in cooperation with the Internet Archive, has digitized its monograph holdings published before 1923. These books are available without charge to anyone with access to the Internet through the Scholar's Portal e-Book platform.

Key Databases: For Statistical Sciences, *MathSciNet* and *Current Index to Statistics*; for Computer Science, *ACM Digital Library* and *INSPEC.*

Special Collections: The Library holds an extensive collection of datasets, including microdata, aggregated statistics and geospatial data covering all disciplines.

Current Gaps: The Library would like to provide collaborative data visualization spaces for use by the U of T community, and would be interested in working with the Departments of Computer Science and Statistical Sciences to consider future options.

Prepared by: Bruce Garrod, Mathematics Librarian and Liaison Librarian, Statistical Sciences; and Ben Walsh, Liaison Librarian for Computer Science; August 2, 2016

Submitted by: Larry Alford, Chief Librarian, University of Toronto Libraries, Date

Appendix D: Student Services Report

Student service information for Quality Assurance Framework
[St. George Campus]

All University of Toronto undergraduate and graduate students have access to student services on all three campuses, Mississauga, St. George (downtown Toronto), and Scarborough, regardless of their 'home campus'. The services and co-curricular educational opportunities provide a complement to the formal curriculum by engaging and challenging students to reach their full potential as learners, leaders and citizens. At Arts & Science in University of Toronto (St. George Campus), these services are organized by Student Life Programs and Services, the Office of the Faculty Registrar, and the students' Colleges. All of these groups support the success of our students from the time they are admitted through degree completion and beyond.

Students have access to comprehensive **physical and mental health care** on campus including a medical clinic, travel medicine services, immunization, contraception and sexual health education. Counselling and treatment options for psychological and emotional concerns include psychotherapy, group therapy and pharmacotherapy, as well as specialized assault counseling services.

Housing needs, including off-campus housing listings and resources for students living independently, are met through the Student Housing Service.

Coaching and education in the development of key **learning skills** – from time management to overcoming exam anxiety – is provided through the Academic Success Centre. The ASC also partners with faculty to integrate success strategies and support into the curriculum.

Students' career exploration and employment services are provided through a **Career Centre** offering resume and interview coaching, workshops, career resources, on and off-campus employment and volunteer listings, job shadowing, and career counseling.

Specialized services are provided for **international students** (orientation, advising, cross-cultural counselling), students with **disabilities** (academic accommodations, advising), students with **children or other family responsibilities** (advising, resources, subsidized child care), **aboriginal students** (academic support, financial counselling) and **lesbian, gay, bisexual and transgender** students (counselling, referrals, equity outreach and engagement).

Participation in **campus life** and **experiential learning** are facilitated through Hart House (clubs, committees, events), the Centre for Community Partnerships (service learning), the Multifaith Centre (interfaith dialogue, events), and the Office of Student Life (leadership development, orientation, recognition and support for student groups, activities.) **Sport and recreational facilities and programs** are provided to all students through both Hart House and the Faculty of Kinesiology and Physical Education.

At Arts & Science, students may also access:
- registrarial services; academic advising, through both Colleges and the Departments of Computer Science and Statistical Sciences
- writing centres
- program-related career services, including advising related to the possibility of a Professional Experience Year

- student activity spaces
- residence life programs and services
- student life programs (orientation; first-year learning communities, etc.)

In addition, academic advising will be available through both the Department of Computer Science (DCS) and the Department of Statistical Sciences (DoSS). Faculty who will regularly teach the Data Science integrative courses will take on an advising role relating to the Data Science Specialist academic program and regarding seeking internships in industry.

The Computer Science Undergraduate Office and the Associate Chair, Undergraduate Studies, DCS, who currently serve all Computer Science students, will provide academic advising to Data Science students on all matters concerning the Computer Science aspects of their studies. The Associate Chair, Undergraduate Studies, DoSS, will advise Data Science students on all matters concerning the Statistics aspects of their studies.

Both the Department of Computer Science and the Department of Statistical Sciences are leaders in developing in-class and co-curricular support for students in their programs. Computer Science provides diverse networking opportunities with alumni, while Statistical Sciences has designed unique initiatives to support international students in first and second year, through a peer-mentoring program.

Appendix E: Appraisal Report

Appraisal Report for New Undergraduate Program Proposal

Undergraduate Specialist in Data Science
Departments of Computer Science and Statistical Sciences
Faculty of Arts and Science
University of Toronto

September 25, 2017

This appraisal report is based on the program proposal provided in advance of our site visit to the University of Toronto (UT), together with information we gathered during our visit on September 22, 2017.

Our visit involved a series of meetings indicated by the attached agenda. These included meetings with the most relevant administrators, with different groups of faculty members from both the Computer Science and Statistical Sciences departments, an enjoyable and informative lunch with a group of 7 students, as well as a valuable period of time in the late afternoon to allow us to formulate the content of our report. The visit was well-organized and provided us with suitable opportunities to collect additional information.

In the report that follows, we provide comments and recommendations on each of the program evaluation criteria listed in the Terms of Reference document provided to us in advance of our visit.

1 Objectives

- *Consistency of the program with the institution's mission and unit's academic plans*

With excellent Departments of Computer Science and Statistical Sciences and the current very strong demand for data scientists in the workplace that is only projected to increase in future, the proposed undergraduate specialist program, as delivered by the two departments, is a natural and desirable addition to those available at UT. The three new integrative/capstone courses, Methods of Data Science I/II/III, make the program unique among those currently available at other Canadian universities. The program seems likely to be very attractive to many students who might otherwise enroll in other programs offered by these two departments.

- *Clarity and appropriateness of the program's requirements and associated learning outcomes in addressing the academic division's undergraduate Degree Level Expectations*

The requirements are very clearly specified in the proposal. The associated learning outcomes are carefully described in the proposal and the roles of each required course in the program in introducing/reinforcing/mastering these learning outcomes are described in detail in the curriculum map provided in appendix H of the proposal. The requirements and the learning outcomes are well-considered and appropriate for an undergraduate program in Data Science.

The proposed program merges a Computer Science concentration in machine learning with selected courses from Statistical Sciences that are most relevant to Data Science. The main addition is three new integrative courses, one in each of years 2, 3 and 4. These new courses

and the program itself are very well designed in terms of the degree level expectations. The program has high expectations of student achievement, but the group of students that we met all seemed capable of meeting such expectations. An appealing aspect of this proposed program is that graduates would be eligible for direct entry into the Ph.D. programs in both Computer Science and Statistical Sciences.

- *Appropriateness of the degree or diploma nomenclature*

The nomenclature seems appropriate.

2 Admission requirements

- *Appropriateness of the program's admission requirements for the learning outcomes established for completion of the program*

The enrollment requirements specify completion of 4.0 FCEs, including minimum grades of at least 70% in CSC148H1, CSC165H1/CSC240H1 and STA130H1, and an average grade in these courses or specified alternatives that meets the program's annual cutoff. On the other hand, the completion requirements specify a total of 3.5 FCEs in first year (1.5 FCE MAT, 1.5 FCE CSC and 0.5FCE STA) indicating a high degree of concentration in first year. We are unclear how this degree of concentration compares to the requirements of other specialist programs at UT, but this almost seems to require that students identify that they wish to enter this program prior to entry into first year. These first year completion requirements will fully prepare the students for the courses required in the second year of the program.

We recommend consideration of paths into the program for students who may not have known they were interested in Data Science in high school. Half of the students we met during our site visit indicated they did not discover their interest in Data Science until their 2nd year or later and we expect that it might take time for certain types of students to realize their true interests. These alternative paths are also likely to increase the diversity of students in the proposed program.

- *Appropriateness of any alternative requirements, if any, for admission into the program such as minimum grade point average or additional languages or portfolios, along with how the program recognizes prior work or learning experience*

Given that this is meant to be a selective program, the 70% minimum grade requirement specified as part of the enrolment requirements seems appropriate. However this requirement will also limit the overall diversity of the program and it may be worth reconsidering the requirement after the program has been established for a few years.

3 Structure

- *Appropriateness of the program's structure and regulations to meet specified program learning outcomes and Degree Level Expectations*

Overall, the program seems well-structured and clearly focused on the specified learning

outcomes. The required existing courses and associated learning outcomes focus on the Knowledge of Methodologies and Application of Knowledge Degree Level Expectations. The three new integrative courses, particularly the capstone integrative course, are critical to addressing the Communication Skills, Awareness of Limits of Knowledge, and Autonomy and Professional Capacity Degree Level Expectations.

Some aspects of the structure and regulations that might benefit from further consideration:

> Elective courses will be relied upon to address the Breadth and Depth of Knowledge Degree Level Expectations. Detailed advising, perhaps including lists of recommended courses or sets of courses, will be required to ensure that students 'investigate how the study of computer science and statistics relates to other academic fields and areas of human creativity', are 'asked to consider how Data Science interacts with other academic fields within the sciences, social sciences, and humanities', and 'choose a specialized program of study that will emphasize more highly developed knowledge and skills' (p.13 of the proposal). We recommend further consideration of how such advising would be implemented as this aspect of the program was not covered in detail in the proposal or during the site visit.

> We wonder why the additional gating mechanism of a minimum grade of 73% in the first integrative course (Methods of Data Science I) is required after entry into the program. Usually students are removed from a program as a consequence of failing its required courses. We recommend removing the additional grade requirement and instead using the standard failure mechanism.

> The proposed program requires 50% more FCEs in Computer Science than in Statistical Sciences. This appears to be a consequence of utilizing existing courses, rather than courses more specifically focused on Data Science, to fulfill the training required by students in the program. During our visit we learned that the machine learning courses in Computer Science are currently being restructured. We recommend that Computer Science consider developing courses that more directly focus on Data Science and that might allow a better balance in the collection of required courses in this program. This might also reduce the number of FCEs required to graduate, giving students more flexibility.

> Several components of importance to Data Science were not apparent in the proposal; examples include data wrangling, data analysis workflow management, approaches to reproducible research, and parallel/distributed/cloud computing, data storage and processing for big data. During our site visit we were reassured to learn that the importance of these components was clearly recognized. We were provided with an earlier version of a more detailed description of the first integrative course, Methods in Data Science I, where several of these components appeared. We also learned that some of the existing required courses in Statistical Sciences are being revised to integrate some of these components, and the planning for how these components would fit into the other two new integrative courses was under active consideration. Additional faculty who work on these systems level issues related to Data Science would also strengthen these aspects of the program. We recommend that detailed

plans for all three of the new integrative courses be developed simultaneously to ensure that these critical courses provide as comprehensive training in Data Science as possible.

4 Program content

- *Ways in which the curriculum addresses the current state of the discipline or area of study*

As clearly described in the proposal, Data Science is an emerging discipline that relies on advanced training in Statistical Science and in those aspects of Computer Science that are most relevant to the analysis of data, particularly of (though not limited to) complex large-scale datasets. The fundamental characteristic of a data scientist is the ability to apply these tools and specialized expertise in an integrated fashion to address real-world problems, and to communicate clearly the findings of their analyses.

Students in this program will receive discipline-specific training through the required existing courses in Mathematics, Computer Science and Statistical Science. But their training as a data scientist will rely very heavily on the three new integrative courses. As described in detail in the proposal, the specific courses required in Computer Science and Statistical Science and these integrative courses distinguish this program as much more than a double major program.

- *Identification of any unique curriculum or program innovations or creative components and their appropriateness*

The three new integrative courses are a unique feature of this program. Other Canadian undergraduate programs in Statistical Science and Computer Science include capstone courses similarly focused on addressing real-world problems, but we are not aware of other programs offering a series of such integrative courses, one in each year of the program. These courses are to be 'team taught by existing faculty in Computer Science and Statistics, who are practicing data scientists', another quite unique and desirable feature. These innovations are not only appropriate but will be essential to the success of this program.

5 Mode of delivery

- *Appropriateness of the proposed mode(s) of delivery (distance learning, compressed part-time, online, mixed-mode or non-standard forms of delivery, flex-time options) to meet the intended program learning outcomes and Degree Level Expectations*

The proposal discusses past innovations by both departments leading to improvements in the learning experience in large-enrolment introductory courses (MOOCs, online learning modules, and online and blended learning, including inverted classrooms). These innovations aim to promote a more active learning environment and students in the program will benefit from any of these innovations that have been integrated into the required existing courses. We heard a little bit about these plans at the site visit, but we recommend further development of the instruction delivery plans for the integrative courses as those are fleshed out.

6 Assessment of teaching and learning

- *Appropriateness of the proposed methods for the assessment of student achievement of the intended program learning outcomes and Degree Level Expectations*

Assessment of students in the existing required courses is in place and presumably appropriate. The three new integrative courses would seem to require specialized assessments of student achievement but the proposal does not describe how these will be carried out. We recommend developing more detailed plans for these courses that include descriptions of how the students will be assessed.

- *Completeness of plans for documenting and demonstrating the level of performance of students, consistent with the academic division's statement of its Degree Level Expectations*

The proposal does not discuss this. We recommend more detailed planning including consideration of how project partners (industrial or academic) might participate in this process, especially for the integrative courses.

7 Resources

- *Adequacy of the administrative unit's planned utilization of existing human, physical and financial resources, and any institutional commitment to supplement those resources to support the program*

Once fully launched, beyond existing courses, the program requires a program director and the delivery of the three new integrative courses. If these were lecture-style courses, given their projected size of 30-40 students, this would be a minor increase in the teaching required of these two relatively large departments. But the focus of these courses on experiential learning and the development of communication skills implies the load associated with delivering one of these courses is likely to be much more than for a lecture-style course. Of course, this depends very much on how these courses are delivered, particularly the experiential learning component.

The proposal does not discuss this directly, but points to current teaching and research stream faculty in both departments with relevant expertise and interests, and to the ongoing recruitment of additional such faculty. It also notes that this proposed program is a priority for a current search for a teaching stream professor in the Department of Statistical Sciences. The proposal seems to be stating that the required faculty resources are already in place. We believe that at the proposed scale (100 students total, 30-40 per year) no additional faculty resources are necessary, but that it is very important to consider how the program will scale as it grows. We believe that the program will be oversubscribed very quickly and if the program does grow, more faculty in both Statistical Sciences and Computer Science will be necessary to ensure the strength of the Data Science program.

The delivery of the three new integrative courses will also require extensive TA support. The

proposal describes this need as 'the usual level of TA support provided by the Faculty of Arts & Science for Computer Science and Statistics courses' but given the special character of these courses it is not clear to us that level of TA support will be sufficient. Again, this depends very much on how these courses are delivered, but in our experience courses focused on experiential learning require significantly more TA support than standard lecture-style classes.

- *Participation of a sufficient number and quality of faculty who are competent to teach and/or supervise in the program*

The proposal provides a list of over 20 'committed faculty'. This is an impressive collection of teaching and research stream faculty. These individuals are capable of providing excellent teaching and supervision in the proposed program.

- *Adequacy of resources to sustain the quality of scholarship and research activities of undergraduate students, including library support, information technology support, and laboratory access*

All the usual resources available at UT will be available to the students in this program. The main potential issue we see is the challenge of delivering high quality experiential learning experiences to the students in each of the three integrative courses. The proposal notes that 'both Departments have developed expertise in running academic programs with internship components, at the undergraduate and the graduate levels, and faculty will be able to support further developments of connections to industry in the context of the Data Science program'. During our site visit we learned that the professional Masters in Applied Computing program now has a Data Science concentration and that the departments plan to utilize the many partnership relationships that have been established in this program to benefit the proposed program. This provides considerable reassurance, but the magnitude of this challenge should not be underestimated, particularly if the program is allowed to grow beyond the scale described in the proposal.

- *Adequacy of and planning for:*
  - *Commitment to provide the necessary resources in step with the implementation of the program*

Not directly addressed in proposal.

  - *Planned/anticipated class sizes*

The proposal provides the projected enrolments until steady-state is expected to be reached with 30-40 students per year in 2023-24. However we believe that demand for this program will be extremely high and that it is worth considering how to scale up the program to accommodate significantly more students.

  - *Provision of supervision of experiential learning opportunities (if required)*

See above comment.

o *The role of adjunct and part-time faculty*

This is not directly addressed in proposal. During our site visit we were told that industrial partners might be invited to help deliver courses and mentor students if and when appropriate.

8 Quality and other indicators

- *Quality of the faculty (e.g., qualifications, research, innovation and scholarly record; appropriateness of collective faculty expertise to contribute substantively to the proposed program)*

Both departments involved are recognized as outstanding, both nationally and internationally, and include a substantial number of excellent faculty members who are data scientists with the appropriate expertise and interests to contribute substantively to the proposed program.

- *Program structure and faculty research that will ensure the intellectual quality of the student experience*

The proposal outlines the strengths of the two departments in both pedagogy and research that are of most direct relevance to this program. The quality of the experiential learning experiences provided in the three integrative courses will largely determine the intellectual quality of the student experience unique to this program (beyond that delivered through the required existing courses). We believe that the key to making these courses effective is to partner with domain experts (industrial and academic) that are working with real data world datasets and to ensure that they are as invested in the courses as the faculty instructors and the students.
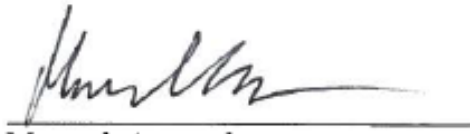
9 Summary

Today, there is very strong demand for data scientists in the workplace, and it is projected to increase for the foreseeable future. The proposed undergraduate specialist program in Data Science is designed to give students the skills they need to work as data scientists and thereby address the widespread demand for such skills. Such a program is a desirable development at UT and we believe it will be very popular.

The proposed program will be jointly delivered by the two departments (Computer Science and Statistical Sciences) that are most directly relevant to this emerging discipline. The excellence of both departments greatly enhances the value of the proposed program. The program takes advantage of many existing courses offered by the two departments and is very well designed in terms of the Degree Level Expectations. The integrative courses are the main new feature of the program and will give students the opportunity to work with real-world data to solve real-world problems. This is one of the most important features of the program. These courses will make this program unique among Data Science-focused programs currently available at other

Canadian universities.

As we outline in the report, we believe that some additional planning is required for the three new integrative courses, Methods of Data Science I/II/III, to ensure that the program functions smoothly and that it can scale as more students learn about the program. We believe that the proposed Data Science specialist program will be very attractive to many students who might otherwise enroll in other programs offered by these two departments, as well as many other students at UT. We have no doubt that this program will provide excellent and challenging training for the students.

Respectfully submitted:

Maneesh Agrawala

John Petkau

UNIVERSITY OF TORONTO
FACULTY OF ARTS & SCIENCE

Department of Computer Science & Department of Statistical Sciences
Specialist in Data Science Program

Agenda for Friday September 22, 2017

Reviewers:
- Professor Maneesh Agrawala, Ph.D., Stanford University
- Professor John Petkau, Ph.D., University of British Columbia

Locations:
- The Dean's Office in the Faculty of Arts and Science is located at Sidney Smith Hall, 100 St. George St., Suite 2005.
- The Department of Statistical Sciences is located at Sidney Smith Hall, 100 St. George Street, Room 6018, Toronto, Ontario, M5S 3G3 (416-978-3452).
- The Department of Computer Science has space in several buildings; the meetings for this visit will be held in the Bahen Centre for Information Technology (BA), 40 St. George Street, Room 4252, Toronto, Ontario, M5S 2E4.

Contacts:
- Teresa Nicoletti, **Tel:** 416-978-5413, **Email:** teresa.nicoletti@utoronto.ca
- Caitlin Burton, **Tel:** 416-978-0445, **Email:** caitlin.burton@utoronto.ca

Time		Location
8:30 to 8:45 am	*Travel time. Reviewers are escorted by Teresa Nicoletti to Sidney Smith Hall.*	
8:45 to 9:30 am	Professor David Cameron, Dean, Faculty of Arts & Science Professor Mary Pugh, Acting Vice-Dean Undergraduate & International	Dean's Office, Sidney Smith Hall (SS), Rm. 2005
9:30 to 10:30 am	Professor Jamie Stafford, Chair, Department of Statistical Sciences Professor Ravin Balakrishnan , Chair, Department of Computer Science	Dept of Statistics, Rm. 6019
10:30 to 10:45 am	*Break*	Dept of Statistics, Rm. 6004
10:45 to 11:45 am	Assistant Professor, Teaching Stream, Nathan Taback, Director Data Science programs, Dept of Statistical Sciences Associate Professor, Teaching Stream, François Pitt, Associate Chair, Undergraduate, Computer Science	Dept of Statistics, Rm. 6004

	Assistant Professor, Teaching Stream, Matt Medland, Director MScAC program, Computer Science	
11:45 to 12:00	*Travel time. François and Matt to accompany reviewers to Bahen Centre for Information Technology*	Dept. of Computer Science, Rm.5250
12:00 to 1:30	**Lunch with undergraduate students** Lebohang Radebe, MScAC student, Data Science Concentration Nidhi Kantak, Stats Loora Zhouran Li, Stats (Robert) Zhi Cheng Xu, Stats Lana Sanyoura, third year in CS/Cog Sci. Erwin-Jeffrey Komguem, third year in Math/CS/Stats. Hisham El-Assi, third year in CS/stats.	Dept. of Computer Science, Rm.5250
1:30 to 2:30	**Computer Science Faculty** University Professor Allan Borodin Assistant Professor Anna Goldenberg Associate Professor, Teaching Stream François Pitt Director MScAC program, Matt Medland Assistant Professor, Roger Grosse Professor Ravin Balakrishnan ,Chair	Dept. of Computer Science, Rm.5250
2:30 to 2:45	*Travel Time. François to accompany reviewers to Department of Statistics.*	
2:45 to 3:30	**Statistical Sciences Faculty** Assistant Professor David Duvenaud Associate Professor, Teaching Stream, Alison Gibbs Professor Keith Knight University Professor Nancy Reid Professor Jeffrey Rosenthal Professor Lei Sun Assistant Proessor, Teaching Stream, Nathan Taback, Director Data Science programs, Dept of Statistical Sciences	Dept of Statistics, Rm. 6004
3:30 to 3:45	*Break and travel time. Nathan to accompany reviewers to Dean's Office.*	
3:45 to 4:45	Time for reviewers alone to review notes, frame the joint report, and prepare any questions for the debrief	Dean's Office, SS, Rm. 2037
4:45 to 5:30 pm	Debrief with Professor Mary Pugh, Acting Vice-Dean Undergraduate & International	Dean's Office, SS, Rm. 2005

Appendix F: Dean's Administrative Response to the Appraisal Report

UNIVERSITY OF TORONTO
FACULTY OF ARTS & SCIENCE

November 9, 2017

Professor Sioban Nelson
Vice-Provost Academic Programs
University of Toronto

RE: Appraisal of the Data Science new undergraduate program proposal

Dear Sioban,

Along with the faculty, staff, and students of the departments of Computer Science and Statistical Sciences, I am very pleased with the external appraisers' positive evaluation of the undergraduate Specialist in Data Science program proposal. I would like to take this opportunity to extend my gratitude to Professors John Petkau and Maneesh Agrawala, for the commitment of their time to this process, as well as their careful appraisal. The Faculty of Arts and Science places a great deal of importance on these appraisals and the care with which the appraisal report was prepared is deeply appreciated.

The appraisers confirmed that the proposed program in Data Science would give students the skills that they need to work as data scientists and would help address the widespread demand for such skills. They identified both the Computer Science Department and the Statistical Sciences Department as excellent departments with an already-existing array of relevant courses upon which the proposed program can be built. They identified the three new courses, Data Science I, II, and III, as one of the most important features of the proposed program; these courses would make the proposed program unique among data-science-focussed programs at other Canadian Universities.

The appraisers identify key issues which we would like to address in the following response.

Section 2. Admission Requirements
The appraisers noted the "high degree of concentration in first year" implied by the program requirements of 3.5 FCEs in first year.

Although this comment was in the "Admission Requirements" portion of the external appraisal, it is a comment on the program structure more than on admission requirements. In the original proposal, the program requirements would have been 3.5 FCE in a student's first year, 4.5 FCE in their second year, 3.0 FCE in their third year, and 2.5 FCE in their fourth and final year. The appraisers flagged this both as an unusually heavy first year's focus but also as a possible deterrent to students who hadn't already discerned an interest in data science in high school. We agree with the appraisers and also noted that the proposed second year program requirements were unusually heavy; heavier than those of the Computer Science Specialist and those of the Statistics Specialist. The program has since been restructured to require only 3.0 FCEs of required preparation for the Data Science Specialist in first year: MAT137Y1/MAT157Y1; MAT223H1/MAT240H1; STA130H1; CSC108H1; and CSC148H1. Also, because students with strong programming backgrounds are able to waive CSC108H1 and enrol directly into CSC148H1, there is an additional possibility to diversify their courses in their first year. The program requirements have been redistributed in the second, third, and fourth years to be 3.5-4 FCE, 3.5 FCE, and 3 FCE respectively, and these changes are now reflected in the proposal.

The appraisers also noted that some students may not realize they want to take the Data Science Specialist until second year.

The program proposal has now been modified to include a route into and through the Data Science Specialist for students who started with a general interest in computer science and later discerned an interest in data science. Similarly, a route is included for students who started with a general interest in statistics and later discerned an interest in data science. This route would be very similar to that taken by any student who started with a general interest in math or in any of the physical sciences. These alternate routes are partly facilitated through the restructuring of the program requirements explained above. They are also facilitated by the University of Toronto having fairly robust first- and second-year offerings during the Summer session; it's not unusual for students to use the summer before their second year to change their trajectory. We feel confident that a student who enters the program through one of these alternate routes would be able to complete it in four years, perhaps with the addition of a summer semester.

Section 3. Structure
The appraisers note that students may require "detailed advising" in selecting their elective courses to maximize their breadth of knowledge.

Such advising on course selection is to be provided by the program director. The program director will have administrative support to more effectively provide this advising. We agree with the appraisers that lists of recommended courses will support students in choosing appropriate electives. To that end, the Faculty of Arts and Science will identify courses that connect the digital/data world to the humanities and social sciences; such courses will be especially interesting and relevant to the Data Science Specialists.

The appraisers recommend removing the requirement of a grade of 73% or higher in JSC270H1 (Methods of Data Science I).

The program proposal has been modified to remove this progression to degree requirement; the program has no progression requirements.

The appraisers note that this interdisciplinary program is weighted notably more heavily toward CSC-designated courses than STA-designated courses.

Most students enter university with a stronger foundation in mathematics or statistics than in computer science. Thus, CSC programs, and introductory CSC courses by extension, are designed for students "starting from scratch" in the field – i.e. with no programming experience. Hence, in order to reach a comparable level of proficiency as they have in Statistics by the end of the program, students need to take a greater number of Computer Science courses. As a result, we have not modified the numbers of CSC or STA courses in this program.

The appraisers recommended that the new joint (JSC) courses be simultaneously developed to ensure thorough training in Data Science, including coverage of topics such as "data wrangling, data analysis workflow management" and others.

The first integrative course for the Data Specialist program (JSC270H1) entered Faculty governance on October 20, and the subsequent two joint courses (JSC370H1 and JSC470H1) will be developed simultaneously in 2018/19. This will allow the departments to draw on experience with the first integrative course when designing the two later-year courses.

2

Section 5. Mode of Delivery

The appraisers heard about past pedagogical innovations by the departments during their visit. They recommend further development of the instructional delivery plans for the integrative courses as those are fleshed out.

The program proposers agree with this comment and will take it to heart when implementing JSC270 in Winter 2019 and while creating JSC370 and JSC470 in Fall 2018 and implementing them in Winters 2020 and 2021. In doing so, they will take advantage of the fact that both the Computer Science and the Statistical Sciences Departments have been active in using new pedagogical methods in their courses. There is significant expertise in their teaching-stream faculty that bears upon online and blended learning, active learning, course modules and so forth.

Section 6. Assessment of Teaching and Learning

The appraisers note that the three new integrative courses (JSC270, 370, and 470H1) "seem to require specialized assessments of student achievement," which were not delineated in the proposal.

The integrative courses will require assessments that focus more on the student's ability to synthesize and apply learning from the other, discipline-specific courses. Also, in the fourth-year integrative course (JSC470) the students will have capstone projects that will involve working with data from industry, the sciences, the social sciences, or the humanities. Students will need to be assessed on their ability to discern what are the questions of interest to the data-holders, how to address those questions, and how to best communicate their findings. To help with effective assessments of this type, the program will rely on local expertise. Both departments have teaching-stream faculty members whose research focuses on pedagogy and student assessment: Jennifer Campbell (CSC), Michelle Craig (CSC). Diane Horton (CSC), Alison Gibbs (STA), Paul Gries (CSC), Nathan Taback (STA), and Bethany White (STA). Professors Gibbs and Taback are directly involved in the Data Science program and will be able to consult with the others.

Further to the above point, the appraisers noted that the proposal did not completely address plans for documenting students' level of performance in line with the division's Degree Level Expectations. Specifically, they recommended planning for how project partners in industry or academia might participate in assessment of students' performance.

Both departments have experience with students working with partners – in industry and within the university. They have access to industrial partners through prior and current ventures such as hackathons and datafests. In addition, the Master of Science in Applied Computing program involves students working on software engineering projects which are supervised by industrial partners. Thus, the departments can leverage relevant experiences to good effect in the Data Science program. In response to the appraisers' comments, section 9 has been updated to include the following; "Available data on specific learning objectives directly related to DLE will be monitored. The program director will monitor course evaluations, course grades, and GPA. As another form of assessment, we will survey data scientists and industry professionals that are part of the integrative courses, interact with students, and see their work first hand. We will also receive qualitative feedback on the program on a continuous basis from these professionals. This will help keep methods and topics in these courses current."

3

Section 7. Resources
The appraisers express two main points here. First, that the nature of delivery of the integrative courses will partially determine how much of a teaching load they present; some delivery styles (e.g. heavily experiential) could present quite a heavy teaching load. Second, and relatedly, demand for this program may be very high in years to come, and – while the presently proposed magnitude does not seem to require additional resources – scaling up the program over time may require more resources (e.g. faculty and TA resources).
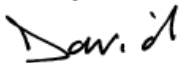
The Department of Statistical Sciences is currently searching for two Assistant Professors, Teaching Stream in the area of "Statistical Computation and Statistical Collaboration". In addition, there is a joint Computer Science/Statistical Sciences search for an Assistant Professor, Tenure-Stream in the area of "Data Visualization". Assuming these searches go well and candidates are hired, these three new faculty members would join us on July 1st, 2018. The current plan is to keep the program fairly small with a steady-state enrolment of about 40 students entering the program each year. This will allow the program to be created and maintained without any additional faculty hiring beyond the current advertised positions. As additional instructor support, there are status-only faculty who are affiliated with the departments who could be asked to guest lecture on occasion.

The appraisers flagged that the interdisciplinary courses JCS270H1, JCS370H1, and JCS470H1 may be especially difficult to find TAs for, partly because of the intensive nature of the courses and partly because of the TAs needing to be well-versed in data science themselves. It is always hard to find TAs for CSC and STA courses because of the high number of students in these courses. This will be less of a problem for the data science program precisely because it will be a small program and will, as a result, require fewer TA hours. And, because graduates of the Data Science Program can apply to the direct-entry PhD programs in CS and in Statistical Sciences, it is quite possible that some of the Data Science Specialists will continue on for PhD studies and would, of course, likely choose to be TAs for the data science undergraduate courses.

In addition, because of increased interest in work-integrated learning, experiential learning, and job readiness, more resources are becoming available in the Faculty of Arts & Science and at the University of Toronto to support such initiatives.

To conclude, we are very grateful to the appraisers for their thoughtful and thorough analysis of the quality of this program, and their recommendations. We believe that the proposed program is novel within Canada, will be of high quality, and that recommendations to maximize the quality of learning experience in the three new integrative courses should be top of mind as this program rolls out to completion.

Sincerely,

David Cameron
Dean and Professor of Political Science

CC: Ravin Balakrishnan, Chair, Department of Computer Science
CC: James Stafford, Chair, Department of Statistical Sciences

4

Appendix G: Vice-Provost's Response to the Dean's Response

UNIVERSITY OF TORONTO

OFFICE OF THE VICE-PROVOST
ACADEMIC PROGRAMS

November 10, 2017

David Cameron
Dean and Professor of Political Science
University of Toronto

Re: Appraisal Report, Proposed New Specialist in Data Science

Dear David,

I am very pleased to receive the appraisal of the proposed Specialist in Data Science which will partially fulfil the requirements for an Honours Bachelor of Science degree. Your administrative response to the appraisal nicely summarizes the report and highlights the specific suggestions made by the appraisers.

The appraisers made specific recommendations around adjusting the program structure to allow for alternative pathways for students who become interested in Data Science later in their program. They also emphasized the importance of student advising so students would be supported when choosing electives. In response, I note that you have accepted many of the appraisers' recommendations, and have, for example, reduced the program requirements in years one and two, outlined alternative pathways in the proposal and, in your administrative response, indicated how student advising will be approached. As well, the proposal and your response now provide more detail on assessment and outline how student performance in the program will be monitored in several ways.

I will be very pleased to recommend this new Specialist program to governance for approval, following approval at the Divisional level.

Sincerely,

Sioban Nelson
Vice-Provost, Academic Programs
cc.

> Teresa Nicoletti, Executive Assistant to the Dean
> Mary Pugh, Acting Vice-Dean, Undergraduate & International
> Tara Casier, Acting Administrative Assistant, Graduate & Undergraduate
> Caitlin Burton, Special Projects, Academic Planning
> Daniella Mallinick, Director, Academic Programs, Planning and Quality Assurance, Office of the Vice-Provost, Academic Programs

Jennifer Francisco, Coordinator, Academic Change, Office of the Vice-Provost, Academic Programs

Appendix H: Curriculum Map

The rows are learning objectives and the columns are the courses where the learning objectives will be met.

I (Introduce) – Key ideas, concepts or skills related to learning outcome are introduced.

R (Reinforce) – Students are developing and becoming increasingly proficient in demonstrating the learning outcome. The learning outcome is reinforced with feedback.

M (Master) – Students demonstrate the learning outcome with high level of independence and a level of understanding and sophistication expected upon graduation.

Table 4: Year 1 Undergraduate Data Science Program

	CSC108H	CSC148H	STA130H	MAT137Y1/ MAT157Y	MAT223H/ MAT240H
Understand basic single and multivariable concepts in calculus				Introduce	
Understand basic linear algebra					Introduce
Understand the fundamentals of probability and statistical theory			Introduce		
Understand the theory and application of statistical inference			Introduce		
Understand the theory and application of machine learning algorithms			Introduce		
Write computer programs	Introduce	Reinforce			
Demonstrate the ability to identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data, and communicating the results			Introduce		
Ability to present the results and limitations of a data analysis at appropriate technical levels for the intended audience			Introduce		
Ethical practice of data analysis and software development			Introduce		
Teamwork skills			Introduce		

Table 5: Year 2 Undergraduate Data Science Program

	STA257H	STA261H	CSC165H/ CSC240H	CSC236H/ CSC240H	CSC207H1	MAT237Y/ MAT257Y	JSC270H
Understand basic single and multivariable concepts in calculus						Reinforce	
Understand the fundamentals of probability and statistical theory	Introduce	Introduce					
Understand the theory and application of statistical inference		Introduce					
Application and critical evaluation of statistical methods and study designs to develop solutions to questions based on real data		Introduce					Introduce
Write computer programs					Reinforce		Reinforce
Familiarity and experience with best practices in software development			Introduce	Introduce	Introduce		
Knowledge of current software infrastructure for handling large data sets							Introduce
familiarity and experience with best practices in software development							Reinforce

	STA257H	STA261H	CSC165H/ CSC240H	CSC236H/ CSC240H	CSC207H1	MAT237Y/ MAT257Y	JSC270H
Understand the fundamental algorithms and data structures needed for storing and processing large amounts of data efficiently			Introduce	Introduce			Reinforce
Understand the limitations of standard data structures and algorithms through fundamental complexity analysis methods			Introduce	Introduce			Reinforce
Develop new data structures and algorithms to support efficient storage and processing of large amounts of data.			Introduce	Introduce			Reinforce
Demonstrate the ability to identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data, and communicating the results							Introduce
Ability to present the results and limitations of a data analysis at appropriate technical levels for the intended audience							Introduce
Ethical practice of data analysis and software development							Reinforce
Teamwork skills							Reinforce

Table 6: Year 3 Undergraduate Data Science Program

	STA302H	STA305H	STA355H	CSC263H/ CSC265H	CSC209H	CSC343H	JSC370H
Write computer programs					Reinforce	Reinforce	
Knowledge of current software infrastructure for handling large data sets					Introduce	Introduce	
Familiarity and experience with best practices in software development				Reinforce	Reinforce	Reinforce	
Understand the fundamental algorithms and data structures needed for storing and processing large amounts of data efficiently				Reinforce	Reinforce	Reinforce	
Understand the limitations of standard data structures and algorithms through fundamental complexity analysis methods				Reinforce			
Develop new data structures and algorithms to support efficient storage and processing of large amounts of data				Reinforce			
Understand the fundamentals of probability and statistical theory	Reinforce	Reinforce	Reinforce				

	STA302H	STA305H	STA355H	CSC263H/ CSC265H	CSC209H	CSC343H	JSC370H
Understand the theory and application of statistical inference			Reinforce				
Application and critical evaluation of statistical methods and study designs to develop solutions to questions based on real data	Reinforce	Reinforce					
Apply machine learning algorithms to solve real world problems							Reinforce
Create appropriate data visualizations for real world problems							Reinforce
Knowledge of current software infrastructure for handling large data sets							Reinforce
Demonstrate the ability to identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data, and communicating the results							Reinforce
Ability to present the results and limitations of a data analysis at appropriate technical levels for the intended audience							Reinforce

	STA302H	STA305H	STA355H	CSC263H/ CSC265H	CSC209H	CSC343H	JSC370H
Ethical practice of data analysis and software development							Reinforce
Teamwork skills							Reinforce

Table 7: Year 4 Undergraduate Data Science Program

	STA414H/ CSC411H	STA442H/ STA437H/ STA457H/ STA480H	CSC421H/ STA303H/STA 347H	CSC373H	Any other 400-level STA	JSC470H
Understand the fundamental algorithms and data structures needed for storing and processing large amounts of data efficiently				Master		
Understand the limitations of standard data structures and algorithms through fundamental complexity analysis methods				Master		

	STA414H/ CSC411H	STA442H/ STA437H/ STA457H/ STA480H	CSC421H/ STA303H/ STA347H	CSC373H	Any other 400-level STA	JSC470H
Develop new data structures and algorithms to support efficient storage and processing of large amounts of data				Master		
Understand the theory and application of statistical inference		Master				
Application and critical evaluation of statistical methods and study designs to develop solutions to questions based on real data.	Master	Master	Master		Master	Master
Apply machine learning algorithms to solve real world problems	Reinforce					Master
Create appropriate data visualizations for real world problems						Master
Write computer programs	Reinforce					Master
Knowledge of current software infrastructure for handling large data sets						Master
familiarity and experience with best practices in software development						Master

	STA414H/ CSC411H	STA442H/ STA437H/ STA457H/ STA480H	CSC421H/ STA303H/ STA347H	CSC373H	Any other 400-level STA	JSC470H
Demonstrate the ability to identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data, and communicating the results						Master
Ability to present the results and limitations of a data analysis at appropriate technical levels for the intended audience						Master
Ethical practice of data analysis and software development						Master
Teamwork skills						Master