Major Program Modification

New Science Minor – Data Science

Nov. 3, 2016

For questions and comments please contact Faisal Qureshi at faisal.qureshi@uoit.ca

1. Introduction

There has been an explosive growth in the amount of data we are able to now create, store and access. The term Big Data has been coined to address the incredible value of intelligent analysis of large volumes of data to support crucial business, financial and strategic decisions. We urgently need individuals capable of turning vast quantities of data, in other words Big Data, into insight. Individuals with this ability, often called *data scientists*, are highly sought after in a variety of industries, from healthcare to finance to business consultancy.

The proposed Data Science minor will enable students enrolled in a diverse set of non-Computer Science (CS) Science-Technology-Engineering-Mathematics (STEM) programs—applied mathematics, physics, chemistry, biology, informatics, etc.—to acquire the computational and data analytics skills needed to be competitive in tomorrow's job market.

The Proposed Minor vs. Other Minors Offered in the Faculty of Science

The Faculty of Science already offers minors in Biology, Chemistry, Mathematics, Physics and Computational Science. Of these, the Computational Science minor is perhaps closest to the proposed Data Science minor. While there is some overlap in course requirements for the existing Computational Science minor and the proposed Data Science minor, there are significant differences in the content covered in two minors and the primary focus of the two minors. The proposed minor, for example, requires students to take courses in databases, information visualization and big data--subjects that are not covered in the existing Computational Science minor. The Computational Science minor aims at developing computer modeling skills for students enrolled in non-CS STEM programs; whereas, the proposed Data Science minor focuses on developing computing and algorithmic skills needed for large scale data analytics in support of scientific discovery and strategic decision making. We suspect that the two minors will attract different groups of students. Consequently, the new minor will have little, if any, impact on the enrollment numbers for the existing Computational Science minor.

2. Degree Requirements

a. Program Learning Outcomes

The minor is designed to provide non-CS STEM students an opportunity to develop knowledge, skills, and tools that will set them for success in the highly lucrative and desirable Data Science job market. We expect that students stand to benefit greatly from the applied computational

and data analytics skills developed by this minor. Specific learning outcomes are:

- Develop a basic understanding of computational tools for data analytics;
- Gain working knowledge over a range of problems, methods and systems for performing data analysis for real-life data sets;
- Attain a good understanding of the applications of data science in a variety of domains; and
- Acquire critical problem solving skills to extract knowledge and understanding from large volumes of data.

b. Admission Requirements

No changes in admission requirements will be made to accommodate the new proposed Data Science minor. Students request to add a minor in any Science discipline.

c. Program Structure

Course requirements for the Data Science minor are listed below:

- CSCI 1040U – Introduction to Programming for Scientists;
- CSCI 2000U – Scientific Data Analysis;
- CSCI 2010U – Principles of Computer Science;
- CSCI 3030U – Database Systems or CSCI 3031U Databases for Scientists[1];
- CSCI 4210U – Information Visualization; and
- CSCI 4030U – Big Data Analytics.

Students who are interested in pursuing additional courses in this area are encouraged, but not required, to take the following courses:

- CSCI 3010U – Simulation and Modeling (pre-req: STAT2010, MATH2072|CSCI2072, CSCI1020|CSCI1030);
- CSCI 4610U – Artificial Intelligence (pre-req: STAT2010, CSCI3070);
- CSCI 4220U – Computer Vision (pre-req: CSCI2010, MATH2050);
- CSCI 4050U – Machine Learning (pre-req: CSCI3070, MATH2050); and
- CSCI 4060U – Multicore and Manycore Programming (pre-req: CSCI3070)

Proposed Calendar Copy

14.13.3 Data Science minor

A minor consisting of 18 credit hours is available in Data Science. This Data Science minor focuses on developing computational and analytical skills needed to turn large volumes of data into insight. Our ability to produce and collect enormous quantities of heterogeneous data is rapidly growing. The term Big Data has been coined to address the incredible value of intelligent analysis of large volumes of data to support crucial business, financial and strategic decisions. We urgently need individuals capable of turning vast quantities of data, in other

[1] This is a new course approved in November 2016. A revised new course template is below, specifying credit exclusions.

words Big Data, into insight. Individuals with this ability, often called *data scientists*, are highly sought after in a variety of industries, from healthcare to finance to business consultancy.

The Data Science minor will enable students enrolled in a diverse set of non-Computer Science STEM programs—applied mathematics, physics, etc.—to acquire the computational and data analytic skills needed to be competitive in tomorrow's job market. The minor will focus on

- developing a basic understanding of computational tools for data analytics;
- gaining working knowledge over a range of problems, methods and systems for performing data analysis for real-life data sets;
- attaining a good understanding of the applications of data science in a variety of domains; and
- acquiring critical problem solving skills to extract knowledge and understanding from large volumes of data.

If the current trends are any indicator, students with a Data Science minor will be highly sought after in whole array of industries.

Course requirements

Students must complete the following courses:

- CSCI 1040U – Introduction to Programming for Scientists;
- CSCI 2000U – Scientific Data Analysis;
- CSCI 2010U – Principles of Computer Science;
- CSCI 3030U – Database Systems or CSCI 3031U Databases for Scientists;
- CSCI 4210U – Information Visualization; and
- CSCI 4030U – Big Data Analytics.

Students who are interested in pursuing additional courses in this area are encouraged, but not required, to take the following courses:

- CSCI 3010U – Simulation and Modeling;
- CSCI 4610U – Artificial Intelligence;
- CSCI 4220U – Computer Vision;
- CSCI 4050U – Machine Learning; and
- CSCI 4060U – Multicore and Manycore Programming.

d. Program Content

Please see Section 2c above.

3. Resource Requirements

a. Faculty Members

Core faculty associated with this specialization include all current Computer Science core faculty (tenure & tenure-track and teaching streams):

- Jeremy Bradbury, *Associate Professor*;
- Christopher Collins, *Associate Professor and CRC in Linguistic Information Visualization*;
- Mark Green, *Professor*;
- Randy Fortier, *Lecturer*;

- Ken Pu, *Associate Professor*;
- Faisal Qureshi, *Associate Professor*; and
- Jaroslaw Szlichta, *Assistant Professor*.

No new core faculty hires are required.

b. Additional Academic and Non-Academic Human Resources

We plan to offer one extra course as a part of this minor. The key difference between the proposed CSCI 3031U (Databases for Scientists) and the existing CSCI 3030U is in the content covered during labs. CSCI 3031U labs are designed to take into account the computer science knowledge and different needs of the non-CS students.

We are of the mind to cross-list the CSCI 3031U lectures with those of CSCI 3030U. Lab sections most likely won't be cross-listed. This will minimize the additional resource requirements for the proposed DS minor.

c. Physical Resource Requirements

Please see Section 3b above.

4. Business Plan

a. Statement of Funding Requirements

The proposed Data Science specialization has no extra funding requirements.

b. Statement of Resource Availability

The proposed specialized as outlined here does not require any new resources.

APPROVAL DATES

Curriculum Committee approval	November 25, 2016
Faculty Council approval	December 14, 2016
CPRC Approval	January 20, 2017
Submission to Academic Council	February 28, 2017

TEMPLATE 8-A

NEW COURSE TEMPLATE

For changes to existing courses see Course Change Template

Faculty: Science		
Course title: Databases for Scientists		
Course number: CSCI 3031U	**Cross-listings:**	____ **Core** x_ **Elective**
Credit weight:	**Contact hours:** 3 Lecture _1.5_ Lab_____Tutorial _____ Other	

CALENDAR DESCRIPTION

The aim of the course is to provide students with an overview of database management systems and tools, an understanding of basic database design and applications, and practical experience of designing and building a relational and other types database. Students will also be exposed to advanced topics such as database implementation, application integration, and query languages in the context of data science and analytics. This course is designed for students with limited programming background. Labs accompanying this course will cover the programming basics needed to understand databases and their applications.

Prerequisites	CSCI 2010U
Co-requisites	
Credit restrictions	CSCI 3030U, SOFE 3700U, INFR 3810U, BUSI 3504U.
Credit exemptions	
Grading scheme	**[x] letter grade [ ] pass/fail**

LEARNING OUTCOMES

1. Understanding of the theory and design of databases.
2. Working knowledge of data analysis and management using query languages.
3. Understanding of the basic implementation principles of database systems.
4. Working knowledge of the applications of database technologies toward scientific data analytics and data science in a broad scope.
5. Understanding data warehousing concepts and OLAP queries reporting.
6. High-level programming needed to interact with databases.

DELIVERY MODE

(check all that may apply) [x] face-to-face [x] hybrid [ ] online

TEACHING AND ASSESSMENT METHODS

Assignments, tests and quizzes either in written, electronic or online format.

The instructor may choose to have a final examination or a final project.

CONSULTATION AND FINANCIAL IMPLICATIONS, WHERE APPROPRIATE

The key difference between the proposed CSCI 3031U (Databases for Scientists) and the existing CSCI 3030U is in the content covered during labs. CSCI 3031U labs are designed to take into account the computer science knowledge and different needs of the non-CS students.

We are of the mind to cross-list the CSCI 3031U lectures with those of CSCI 3030U. Lab sections most likely won't be cross-listed.

APPROVAL DATES

Date of submission	November 2016
Curriculum Committee approval	November 2016
Faculty Council approval	November 2016