BS Data Science Curriculum

Version 1.0, 29 August 2019

1 Executive Summary

This document describes the courses and sequence in which said courses are to be taken to obtain a BS in data science at the University of Oregon.

2 Introduction

A fundamental aspect of the UO Data Science Program is the set of degrees, both undergraduate and graduate, delivered by the program. Many of the UO's comparator institutions have created MS programs; discussion of graduate programs in data science will be the subject of another report. This report focuses on the undergraduate data science program.

One aspect of the charge to the Data Science Visioning Committee was to recommend the educational framework for the initiative. In the discussions of UO undergraduate offerings in data science[1], there was general consensus that the most meaningful form of bachelor's degree would be one in which general data science principles were covered side-by-side with domain-specific applications of the general approaches; a heavy admixture of domain-specific considerations is fundamental.

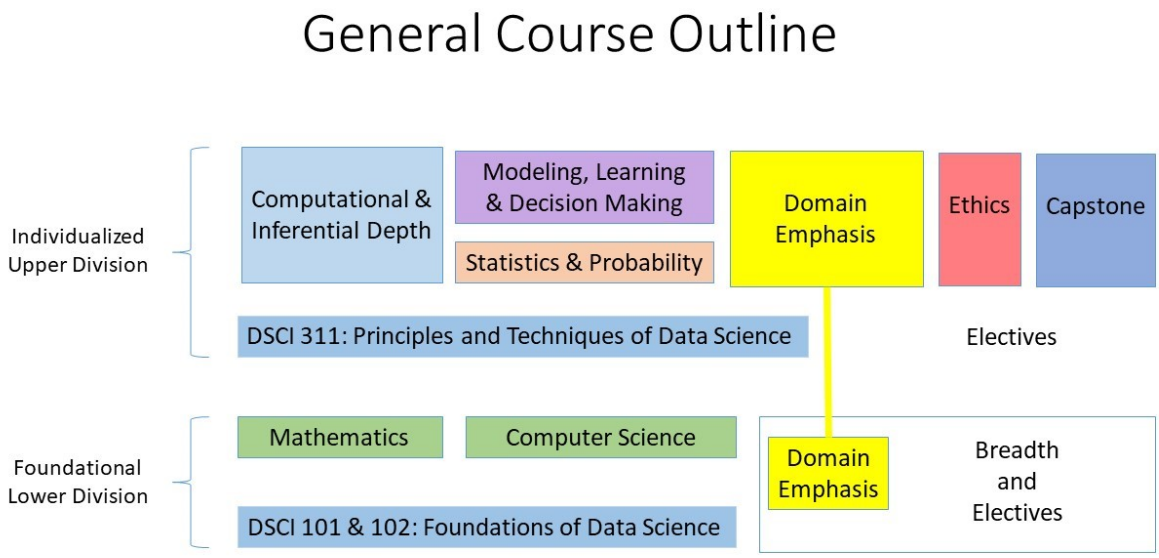
A review of online materials from other AAU institutions, plus Oregon State and Portland State, was conducted in order to understand what, if any, undergraduate programs were offered by each institution[2]. Undergraduate programs were then compared with the above consensus; alignment with that consensus triggered a more in-depth review of four programs: Iowa State, UC Berkeley, UC San Diego, and University of Rochester.

The Berkeley curriculum, with its support for 25 domain emphases, appears to have the strongest and widest mixture of theory and domain emphasis; additionally, it has a framework across lower-division and upper-division courses in several categories that shows substantial thinking about how to best prepare these students for future success. This report documents our adaptation of the Berkeley curriculum to the University of Oregon.

[1] See "A Vision for Implementing the Presidential Initiative in Data Science at the University of Oregon", https://datascience.uoregon.edu/sites/datascience1.uoregon.edu/files/data_science_report_2018.pdf.

[2] See "A Review of BS Data Science Programs".

3 The UO BS in Data Science Curriculum



The above figure shows the general course outline for a student pursuing a BS in Data Science with a domain emphasis. Each of the boxes are described in additional detail in the following sections. An additional domain specialization course may be taken in lieu of the capstone senior project.

3.1 Required courses in Computer Science

CIS 210-212 cover computational thinking, object-oriented programming, and beginning data structures. Thus, students must take 12 credit hours of computer science.

3.2 Required courses in Mathematics

MATH 251-252 cover differential and integral calculus. MATH 341-342 linear algebra. Thus, students must take 16 credit hours of mathematics.

3.3 DSCI 101 & 102

Two new courses, DSCI 101-102, will be developed to provide a broad, but shallow, introduction to data science. DSCI 101 is a core education course, with no pre-requisites. DSCI 102 has DSCI 101 as a pre-requisite. Thus, students must take 8 credit hours for these required courses.

3.4 DSCI 311

One new course, DSCI 311, will be developed to cover three areas in depth: handling data in Python, including Pandas, data Cleaning, EDA, and handling text; Dimensionality and PCA; and Regression, Classification, and Inference. Thus, students must take 4 credit hours for this required course.

3.5 Probability and Statistics

One new course, DSCI/MATH 345M: Probability and Statistics for Data Science, will be developed by CIS and MATH. Thus, students must take 4 credit hours for this required course.

3.6 Computational and Inferential Depth

CIS offers many courses from which students can choose in this area. Students must take 3 courses from this list, totaling 12 credit hours.

Examples include:

- Intermediate Data Structures
- Computer Organization
- Introduction to Software Engineering
- Applied Cryptography
- Intro to Databases
- Intermediate Algorithms
- C/C++ & Unix
- Visualization
- Operating Systems
- Intro to Internet
- Distributed Systems and Network Security
- Programming Languages
- Software Engineering
- Intro to AI

Other courses from MATH and other departments may be added to this list.

3.7 Modeling, Learning, and Decision Making

A new course, DSCI/CIS 372, entitled "Machine Learning for Data Science", is being developed. Students must take this course, worth 4 credit hours.

3.8 Human Contexts and Ethics

The Philosophy department offers a course in technology ethics. Students are required to take this course, worth 4 credit hours.

3.9 Domain Emphases

The domain emphasis consists of completing 2-3 courses (8-12 credits) in the domain core, followed by a minimum of 3 courses (12 credits) of domain specialization. For each domain emphasis, a curated list of courses has been developed for both the core and specialization component.

Thus, students will cover 20-24 credit hours in their chosen domain.

The initial, proposed domain emphases are:

3.9.1 Biology

3.9.1.1 Domain Core – select at least two

- BI 211 General Biology I - Cells 4 cr
- BI 212 General Biology II - Organisms 4 cr
- BI 213 General Biology III - Populations 4 cr

3.9.1.2 Domain Specialization – select at least three

- BI 360 Neurobiology 4 cr
- BI 485 Techniques in Computational Neuroscience 4 cr
- BI 320 Molecular Genetics 4 cr
- BI 493 Genomic Approaches and Analysis 4 cr
- BI 370 Ecology 5 cr
- BI 471 Population Ecology 4 cr

3.9.2 Geography

3.9.2.1 Domain Core – take all three

- GEOG 181 Our Digital Earth 4 cr
- GEOG 281 The World & Big Data 4cr
- GEOG 481 GIScience I 4 cr (Formerly 200-level)

3.9.2.2 Domain Specialization – select at least three

- GEOG 482 GIScience II 4 cr
- GEOG 485 Remote Sensing I 4cr
- GEOG 486 Remote Sensing II 4cr
- GEOG 490 Special Topics 4cr (must have approval from the DSCI director of undergraduate studies)
- GEOG 491 Advanced GIS: Python 4cr
- GEOG 493 Advanced Cartography 4cr
- GEOG 496 Location-Aware Systems 4cr
- GEOG 498 Geospatial Project Design 4cr

3.9.3 Accounting Analytics

3.9.3.1 Domain Core – take all three

- BA 101 Introduction to Business 4 cr
- BA 215 Accounting: Language of Business Decisions 4 cr
- EC 201 Introduction to Economic Analysis: Microeconomics 4 cr

3.9.3.2 Domain Specialization (assumes a 4th course fulfills the DS capstone requirement, the ACTG 410 capstone below is designed to have a major capstone assignment)

- Required:
    - BA 240[3]
    - ACTG 350[4] Intermediate Accounting I 4 cr

  Take two of these three:

    - ACTG 410[2] Accounting Data & Analytics 4 cr
    - ACTG 410 Accounting Data & Analytics Capstone 4 cr
    - OBA 410[5] Predictive Analytics 4 cr

3.9.4 Marketing Analytics

3.9.4.1 Domain Core – take all three

- BA 101 Introduction to Business 4 cr
- BA 215 Accounting: Language of Business Decisions 4 cr
- EC 201 Introduction to Economic Analysis: Microeconomics 4 cr

[3] Students who already have significant experience with Excel can test out of BA 240.
[4] We are prepared to drop EC 202 as a pre-requisite, and alter the pre-requisites to accept "ACTG 213 or BA 215" and "MATH 243 or DSCI/MATH 345"
[5] We are prepared to alter the pre-requisite to accept "OBA 312 or DSCI/MATH 345" and "OBA 335 or DSCI 311"

3.9.4.2 Domain Specialization (assumes a 4th course fulfills the DS capstone requirement)

Required:

- BA 240[1]
- BA 317 Marketing: Value for Customers 4 cr
- MKTG 390[6] Marketing Research 4 cr

Take one of the following two:

- MKTG 395 Marketing Analytics 4 cr
- OBA 410[3] Sports Analytics 4 cr

3.9.5 Linguistics

3.9.5.1 Domain Core – take both

- LING 301 Linguistic Analysis 4 cr
- LING 302 Linguistic Behavior 4 cr

3.9.5.2 Domain Specialization – take at least three

- LING 435 Morphology & Syntax 4cr
- LING 451 Functional Syntax I 4cr
- LING 452 Functional Syntax II 4cr
- LING 493 Corpus Linguistics 4cr
- CIS 410nlp Natural Language Processing 4cr

3.10 Capstone Senior Project

The final 4 credits for the degree can be earned in one of two ways:

1. an additional course from the domain specialization list can be taken for credit; or
2. if a student has a Data Science GPA >= 3.75 AND their emphasis domain has elected to co-advise Capstone projects, the student may take DSCI 411, in which the student applies the techniques and skills acquired in the program to demonstrate domain-specific insight from domain-specific raw data The project will be co-advised by an individual from the DS program and an individual from the domain. The outcome of the project is a report and computational artifacts.

[6] BA 317 is an allowable substitute for MKTG 311

4 Summary

The total SCH load for a student pursuing a BS in Data Science with an emphasis in X is shown in the following table.

Area	Hours
Data science core	12
Mathematics	16
Computer Science	12
Domain core	8-12
Computational and Inferential Depth	12
Probability	4
Modeling, Learning, and Decision Making	4
Human Contexts and Ethics	4
Domain specialization	12
Capstone project (or additional domain specialization course)	4
TOTAL	88-92

BS Flight Path

	Fall	Winter	Spring
Year 1	CIS 210	CIS 211 DSCI 101	CIS 212 DSCI 102 MATH 251
Year 2	Dom[a] 1 MATH 252	Dom 2 MATH 341	[Dom 3] MATH 342
Year 3	DSCI 311 DSCI/MATH 345	DSCI/CIS 372 PHIL 423	CID[b] 1
Year 4	Dom 4 CID 2	Dom 5 CID 3	Dom 6 DSCI 411 (or Dom 7)

[a]Dom = domain emphasis
[b]CID = Computational and Inferential Depth
N.B. an additional course from the curated domain specialization list may be taken in place of the capstone senior project.