

# Computing Across the Disciplines: Study Programs in Computational and Data Sciences at the University of Oslo

A comprehensive path from undergraduate studies to graduate studies (Master and PhD levels) in Computational and Data Sciences at the University of Oslo

Working group members: Andreas Austeng (Informatics), Morten Hjorth-Jensen (Physics), Anders Maltre-Sørensen (Physics), Geir Kjetil Sandve (Informatics), Arne Bernhard Sletsjøe (Mathematics), and Geir Olve Storvik (Mathematics)

## Executive Summary

We propose to establish, with startup Fall semester 2023, a new Bachelor of Science program in Computational and Data Sciences at the University of Oslo. The new undergraduate program is meant to be a multidisciplinary program with a strong methodological base of courses in Mathematics, Statistics, Computational Science and Data Science, with a suggested number of ECTS of 90.

The remaining credits, with elective courses, will allow students to specialize in different disciplines, like Bioscience, Chemistry, Geoscience, Informatics, Physics, Mathematics, Statistics or other physical sciences. The program's multidisciplinary nature entails thus a close collaboration between departments at the Faculty of Mathematics and Natural Sciences.

We propose also that the two existing Master of Science programs, Computational Science and Data Science, merge into one Master of Science program with the same name as the undergraduate program. The new Master of Science program will include all the study directions and study specializations from the previous two Master of Science programs. Planned startup for the new program is also Fall 2023.

Combined with the newly started **CompSci** PhD program, our proposal is meant to enhance the visibility of Computational Science and Data Science at the University of Oslo, from undergraduate studies to graduate studies (Master of Science and PhD levels). It will also align well with many of the new research and educational initiatives on Computational Science and Data Science at the University of Oslo, in particular the newly established **dScience** center.

## Suggested Milestones

If the Faculty of Mathematics and Natural Sciences at the University of Oslo decides to establish a Bachelor of Science program in Computational and Data Sciences, we propose to establish a working group with the mandate to develop a curriculum and program structure with proposed study specializations, by December 15 2021. We would like to propose the following milestones:

- Establish working group by July 1, 2021
- Structure of BS program Computational and Data Sciences with courses and study specializations by December 15, 2021
- Finalization of program proposal by March 15, 2022
- Finalization of program details by October 31, 2022.
- Admission applications ready for start fall semester 2023
- The same working group can also discuss the merging of the Master of Science programs Computational Science and Data Science into one program.
- Finalize structure of new Master of Science program Computational and Data Sciences by December 15, 2021.
- Final approval of new Master of Science program by March 15, 2022
- Finalization of MSc program details by October 31, 2022.
- Admission applications for MSc program ready for start fall semester 2023

## Introduction and Background

Computational Science and Data Science play a central role in scientific investigations and are central to innovation in most domains of our lives. These fields underpin the majority of today's technological, economic and societal feats. We have entered an era in which huge amounts of data offer enormous opportunities, but only to those who are able to harness them. The 3rd Industrial Revolution will alter significantly the demands on the workforce. To adapt a highly-qualified workforce to upcoming challenges requires strong fundamental bases in STEM fields. Computational Science and Data Science can provide such bases at all stages. Most of our students at both the undergraduate and the graduate level are unprepared to use computational modeling, data science, and high performance computing – skills valued by a very broad range of employers. The above developments, needs and future challenges, as well as the exciting developments which are now taking place within quantum computing, quantum information and data driven discoveries (data analysis and machine learning) will play an essential role in shaping future technological developments. Most of

these developments require true cross-disciplinary approaches. In order to meet many of these challenges we propose here that the Faculty of Mathematics and Natural Sciences at the University of Oslo, establishes, with start Fall 2023, a new undergraduate program in Computational and Data Sciences.

Furthermore, we propose also that the two Master of Science programs, Computational Science and Data Science merge into one program. This will enhance the visibility of the two fields and most likely attract many new students. Combined with the newly established **CompSci** PhD program, this has the potential to offer new educational and research paths in computational science and data science at the University of Oslo, with ramifications to other colleges like the Humanities, the Social Sciences, Medicine, and other. These educational paths can lay the foundation for a true multidisciplinary education in Computational and Data Sciences at the University of Oslo. These study programs have the potential to serve all disciplines and need not be confined to Mathematics and the Natural Sciences. The long-term overarching aim is to prepare our students for the technological progress and scientific developments of the 21st Century.

## Brief overview of Computational and Data Sciences

Data science focuses on the development of tools designed to find trends within data sets that help scientists who are challenged with massive amounts of data to assess key relations within those data sets. These key relations provide hooks that allow scientists to identify models which, in turn, facilitate making accurate predictions in complex systems. For example, a key data science goal on the biological side would be better care for patients (e.g., personalized medicine). Given a patient's genetic makeup, the proper data-driven model would identify the most effective treatment for that patient.

Computational science focuses on the development of predictive computer models of the world around us. As study of physical phenomena through experimentation has become impossible, impractical and/or expensive, computational modeling has become the primary tool for understanding—equal in stature to analysis and experiment. Although we can now design an entire commercial aircraft through simulation alone (e.g., the Boeing 777), there are many fundamental problems in science and engineering that are beyond the scope of modern computers with current computational methods. The discipline of computational science deals with the development of new methods that make challenging problems tractable on modern computing platforms, providing scientists and engineers with key windows into the world around us.

The University of Oslo has a multitude of exciting research directions in Computational and Data Sciences, spanning from basic research in Artificial Intelligence and Machine Learning to applications in a great variety of disciplines. Many important recent advances in our understanding of the physical world have been driven by large-scale computational modeling and data analysis, for

example, the 2012 discovery of the Higgs boson, the 2013 Nobel Prize in chemistry for computational modeling of molecules, and the 2016 discovery of gravitational waves.

The new Bachelor of Science program will give students a formal methodological basis in Mathematics, Statistics, Computational Science and Data Science and prepare them for further studies in these fields as well as other disciplines where such competences and knowledge play a central role.

## **Educational programs in Computational and Data Sciences**

The University of Oslo offers two Master of Science programs, one on Computational Science with ten study directions and one on Data Science with five study specializations. These programs and their eventual merge, together with a new Bachelor on Computational and Data Sciences have the potential to form the basis for educational efforts across disciplines and colleges (Faculties) at the University of Oslo. They offer the possibility to strengthen the digital skills and competences of our students through the development of a comprehensive set of courses from undergraduate to graduate levels. The courses and degree programs will give students across the university exposure to practical computational methods, understanding how to analyze data and more generally to the idea of computers as problem-solving tools.

### **Bachelor of Science in Computational and Data Sciences**

The new undergraduate program in Computational and Data Sciences will offer a basic training in central elements in Mathematics, Computational Science, Data Science, Statistics and Informatics. The core part of the program is made up of central courses (see our description proposal in the appendix) in the above fields. Most of these courses are already available, while some courses will have to be developed in close collaboration with the involved departments at the Faculty for Mathematics and Natural Science. The suggested core block of courses is set to 90 ECTS and will be compulsory for all students entering the program. These courses will also be the main requirements for admission to all study directions in the new Computational and Data Sciences Master of Science program.

The remaining 80 ECTS are elective and students can select from a large range of disciplines such as Mathematics, Statistics, Physics, Chemistry, Geoscience, Bioscience, Astrophysics and Informatics. Ten ECTS are reserved for Examen Philosophicum. With the core block of courses combined with recommended combinations of additional courses, students should be able to apply for most of the Master of Science programs at the Faculty for Mathematics and Natural Science.

The suggested compulsory courses are listed in the appendix.

## **Master of Science in Computational and Data Sciences**

Presently the Faculty of Mathematics and Natural Sciences at the University of Oslo offers two Master of Science programs; one in Computational Science (CS) with ten study directions and one in Data Science (DS) with five specializations. The CS program includes almost all disciplines at the Faculty of Mathematics and Natural Sciences. Seven departments are involved through ten study directions in Astrophysics, Bioinformatics, Bioscience, Chemistry, Geoscience, Imaging and Biomedical Computing, Materials Science, Mathematics and Mechanics and Physics. These study directions cover both applications and methodological developments. The DS Master of Science program was established via a collaboration between the Departments of Informatics and Mathematics and has its focus on developments of methods pertinent to Data Science, Mathematics and Statistics. The program offers five specializations, Statistics and Machine Learning, Database Integration and Semantic Web, Data Science and Life Science, Language Technology and Digital Image Processing.

In order to enhance the visibility of Computational Science and Data Science we propose to merge these two programs into one program, maintaining the study directions and specializations. Within the new program we propose two methodological directions towards computational science and data science, respectively in addition to a more applied direction, combining both computational science and data science tools towards other fields. Compared to the existing master program in Computational Science, the new program will have a stronger focus on methodology while at the same time keeping the existing 10 study directions within the applied direction. Compared to the existing master program in Data Science, the methodological focus is kept but with the additional possibility of more applied directions.

The Bachelor and Master of Science programs will most likely increase the visibility of Computational and Data Science at the University of Oslo, and hopefully lead to a larger recruitment of interested students.

## **Appendix: Outline of degree programs and courses**

### **Suggested Themes and Courses for a Computational and Data Sciences Bachelor of Science program**

Here we present a list of courses and themes which could enter the common (compulsory) part of the Computational and Data Sciences bachelor program. Note that what follows is a mere suggestion from our side. If the establishment of such a Bachelor of Science program is approved, a dedicated working group will finalize a more detailed proposal by December 15, 2021.

Courses marked in red are suggested as compulsory. Several of these courses are existing ones while other courses need to be developed or modified.

In some of the paragraphs below, we have included topics to be discussed when outlining the final program structure.

**Mathematics** We suggest that the central topics and themes covered by the existing mathematics courses listed here, are incorporated as compulsory courses. They correspond to 30 ECTS in total and are

1. MAT1100, 1st year
2. MAT1110, 1st year
3. MAT1120, 2nd year

In addition to the compulsory courses many of the students in the program will need a deeper basis in mathematics.

Optimization, discrete mathematics and differential equations are all fundamental areas in mathematics, with huge importance for data science and the natural sciences. Optimization deals with minimizing or maximizing functions, usually subject to constraints, and many problems in statistics, machine learning etc are of this type. Partial differential equations are the tools for modeling and solving dynamic phenomena in natural science. Discrete mathematics contains the mathematics of discrete sets and structures, and is useful for search, discretizations, combinatorial analysis etc.

Suggested senior (last three semesters) undergraduate courses are:

1. MAT3100 Linear optimization
2. MAT3360 Introduction to partial differential equations
3. MAT2250 Discrete mathematics

There are other mathematics course on the borderline bachelor/master that are relevant for this program. In particular, we plan a new course:

MAT 4000xy Advanced linear algebra and applications

Here advanced topics and methods in matrix theory and graph theory are combined. One application treated is clustering (a main topic in machine learning/statistics), where spectral clustering is discussed mathematically and computationally.

**Programming and computer science** Programming is central to any form of computational and data science and is included with 20 compulsory ECTS in the proposed program. The form of programming needed by CS and DS students is less obvious. At least today, many CS and DS practitioners and/or researchers mostly write small scripts, typically consisting of a straightforward sequence of numeric computations. A central question for the program is whether and to what degree (all or a portion of) CS and DS students should gain competence on how to relate constructively to the use of data/computational science in larger/more complex software systems and whether they should have competence on how to analyze and improve computational efficiency.

A question we have left open for now is whether the course variants IN1900/IN1910 should be compulsory for all students. This would ensure an identical programming foundation for all students that later courses could build on, or whether students with stronger interests in the direction of informatics (or integration of CS/DS in larger systems) should be allowed to alternatively choose the course combination IN1000/IN1010. Specifically, a challenge with allowing the IN1000/IN1010 combination would be that such students would gain very limited practical experience with libraries for numerical analysis, which might not work well together with the remaining course combination.

A second question is whether CS and DS students should be expected to have one compulsory or recommended course in an informatics direction, beyond the two programming courses. An introductory course in algorithms could be such a candidate.

Students that have an interest in strengthening their competence in the informatics direction would have a range of courses at the Department of Informatics that could be elected, including: IN2010, IN2040, IN2080, IN2090, IN2100, IN3020, IN3030, IN3040, IN3110, IN3130 and IN3240.

1. **IN1900** 1st year, 10 ECTS (alternatively allow IN1000 to be chosen).
2. **IN1910** (alternatively allow IN1010 to be chosen), 10 ECTS
3. IN2010, IN2040, IN2080, IN2090, IN2100, IN3020, IN3030, IN3040, IN3110, IN3130 and IN3240, elective. 10 ECTS for each course.

**Computational Science/Numerical Methods** We propose that the Computational Science axis of the program is represented by courses in Numerical Mathematics plus elective courses in for example Computational Physics and similar other more senior undergraduate courses. We propose that the Computational Science axis has 20 ECTS which are compulsory plus 10 ECTS which are elective. The suggested courses are

1. **CS I** Could be revised version of MAT-INF1100, first year
2. **CS II**, second year, could be MAT3110
3. Numerical Methods III, elective, third year, FYS3150, Computational Physics, third year and other alternatives.

**Probability and Statistics** 20 ECTS + 10 ECTS elective (also smaller 5 ECTS modules). Here we can think of courses tailored to specific disciplines as well.

In order to understand methods within data science (or perhaps here restricted to data analysis), uncertainty, random variables and probability are important concepts that needs to be learning. For most students these are new concepts that they use time to learn properly.

1. **Introduction to probability and statistics DS I**, first year, preferentially a revised version of STK1100.

An introductory course (10 ECTS) should introduce probability theory (including conditional probabilities and Bayes theorem, some common distributions, expectations, variances, covariances), statistical inference (the likelihood function, parametric modeling, Maximum Likelihood estimation, bias in estimation, confidence intervals, Bootstrapping and resampling, some simple models). In order to make the course attractive for several bachelor programs, the course should/could include relevant examples/case studies from other disciplines.

2. **Statistical learning and machine learning, DS II**, 2nd year, preferentially a revised version of STK2100.

A second course (10 ECTS) should focus on statistical/machine learning methods, starting with linear models (going a bit in depth on this) and then demonstrating *some* advanced methods (e.g nearest neighbor, neural network). Possibly also examples of unsupervised learning/data mining.

3. Data Analysis, AI and Statistics III, third year

Possible courses here are STK3100 - Generalized linear models, IN3050 – Introduction to Artificial Intelligence and Machine Learning, FYS-STK3155 – Applied Data Analysis and Machine Learning. All these courses correspond to 10 ECTS.

The Department of Mathematics (section for statistics and Data Science) is currently looking at modifications of STK1100 and STK2100 following the ideas above. A main change from the current set of courses is that STK2100 can be taken with only STK1100 as background (today it also is built on STK1110).

**Discipline oriented courses** These can be specific courses that lead to the requirements of a bachelor degree in life science, physics, chemistry etc. Most bachelor programs have a requirement of 60-90 ECTS within the specific fields. As an example, the Physics and Astronomy bachelor requires 60 ECTS in mathematics and programming and 70 ECTS in physics courses, plus 10 ECTS for Examen philosophicum and 40 elective ECTS.

**Program structure** We present a possible program structure with elective courses (Examen philosophicum, which corresponds to 10 ECTS, is a compulsory course not included in the recommended block of 90 ECTS).

The example here has a compulsory CS+DS block of 90 ECTS. Adding 10 ECTS for EX-PHIL, this leads to a total of 100 ECTS.



Table 1: Basic Structure of the program

First Semester	MAT1100	IN1900	CS I
Second Semester	MAT1110	DS I	Elective
Third Semester	MAT1120	IN1910	Elective
Fourth Semester	CS II	DS II	Elective
Fifth Semester	Elective/EX-PHIL	Elective	Elective
Sixth Semester	Elective/EX-PHIL	Elective	Elective
ECTS	10	10	10

The remaining courses will be described in more detail when various study specializations are developed.