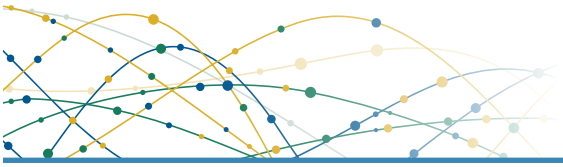


Consensus Study Report

May 2018

HIGHLIGHTS

Data Science for Undergraduates:
Opportunities and Options

As our economy, society, and daily life become increasingly dependent on data, work across nearly all fields is becoming more data driven, affecting both the jobs that are available and the skills that are required. At the request of the National Science Foundation, the National Academies of Sciences, Engineering, and Medicine were asked to set forth a vision for the emerging discipline of data science at the undergraduate level. The study committee considered the core principles and skills undergraduates should learn and discussed the pedagogical issues that must be addressed to build effective data science education programs. *Data Science for Undergraduates: Opportunities and Options* underscores the importance of preparing undergraduates for a data-enabled world and recommends that academic institutions and other stakeholders take steps to meet the evolving data science needs of students.

DATA SCIENCE PROGRAMS AND PATHWAYS

Today, the term "data scientist" typically describes a knowledge worker who is principally occupied with analyzing complex and massive data resources. However, the field of data science as considered in this report spans a broader array of activities including data collection, storage, integration, analysis, inference, communication, and ethics. ***The modern workforce requires both a large population with a basic understanding of data science and a specialized cadre of graduates with highly developed data science skills acquired through substantial coursework and practice.***

RECOMMENDATION: To prepare their graduates for this new data-driven era, academic institutions should encourage the development of a basic understanding of data science in all undergraduates.

RECOMMENDATION: Academic institutions should embrace data science as a vital new field that requires specifically tailored instruction delivered through majors and minors in data science as well as the development of a cadre of faculty equipped to teach in this new field.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The new data science programs will initially combine elements from existing courses, such as those in computer science, statistics, business analytics, information technology, optimization, applied mathematics, and numerical computing. The organization and scope of these programs will vary depending on the culture of a given institution and the aims of its students. Over time, as features of the new data-driven era take shape, academic programs will be compelled to develop new skill clusters, and a body of distinctive courses and instructional materials will emerge.

RECOMMENDATION: As data science programs develop, they should focus on attracting students with varied backgrounds and degrees of preparation and preparing them for success in a variety of careers.

Graduates of these programs will work in virtually every job sector and will serve in a number of roles, including operating the systems on which analyses are run, preparing data for analysis, defining and coordinating analysis, visualizing information, and supporting data-driven decision making to uncover the stories buried in the data. Others who use data science skills will be journalists, administrators, artists, lawyers, teachers, and other workers who need some ability to understand and use data. ***A wide variety of instructional programs will be needed to prepare students for the data-enriched world of the future.***

RECOMMENDATION: Academic institutions should provide and evolve a range of educational pathways to prepare students for an array of data science roles in the workplace.

Academic institutions are stepping up to data science educational challenges with a variety of programs and educational pathways. Several 4-year undergraduate institutions offer data science majors and minors—serving not only those students pursuing data science as a career but also those students who want to acquire data skills while majoring in another field. Two-year institutions are starting to introduce associate degrees and certificates in data science to prepare students to transfer to 4-year programs or to give them skills to compete in the workforce. Summer programs enable undergraduate students to build up data science skills rapidly. Boot camps and intensive training programs help workers pursuing continuing education to acquire and refresh the skills needed to join the growing data science workforce. Massive open online courses in data science serve as points of entry for all kinds of students and provide flexible professional development opportunities for instructors.

These trailblazers in data science show what is possible, but there are significant challenges hampering the development of data science programs more broadly and pervasively. For example, there is a shortage of faculty in this rapidly evolving area. Enlisting and training existing faculty will be essential in the short term, and recruiting new faculty will be important in the long term. Institutions may need to consider how to create incentives for faculty from multiple departments and fields to collaborate to develop and deliver curricula that best meets students' needs. At an institutional level, new data science courses and programs will affect enrollment, budgets, classroom allocation, computing resources, and scheduling. These challenges, among others, will need to be addressed to ensure the success of undergraduate data science students.

KEY PRINCIPLES FOR DATA SCIENCE STUDENTS

Regardless of the type of program, certain key principles need to be covered to give all students the ability to make good judgments, use tools responsibly and effectively, and ultimately make good decisions using data. ***Students will need training in data acquisition, modeling, management and curation, data visualization, workflow and reproducibility, communication and teamwork, domain-specific considerations, and ethical problem solving.*** They will need exposure to material from multiple disciplines—notably, mathematics, statistics, and computer science.

Ethical considerations should also be central to any data science education effort so that students learn to recognize ethical issues and apply a high ethical standard. Students should be well versed in the proper methods for deciding what data to collect, obtaining permissions to use data, crediting the sources of data, validating the data's accuracy, minimizing bias, safeguarding the privacy of individuals referenced in the data, and using the data correctly without inappropriate alteration.

RECOMMENDATION: Ethics is a topic that, given the nature of data science, students should learn and practice throughout their education. Academic institutions should ensure that ethics is woven into the data science curriculum from the beginning and throughout.

RECOMMENDATION: The data science community should adopt a code of ethics; such a code should be affirmed by members of professional societies, included in professional development programs and

curricula, and conveyed through educational programs. The code should be reevaluated often in light of new developments.

EVOLUTION AND EVALUATION

The evolution of data science programs will be affected by a broad range of factors, including their initial home and structure, the needs and interests of students, and institutional culture. Although new programs could be launched by combining existing courses and materials, over time new classes and materials will need to be developed. Institutions will need to think through the pathways students are taking into data science and how to create bridges and remove barriers. Academic and career advising will be vital parts of data science programs; the advising programs will themselves need to evolve as the field and the market for graduates mature.

RECOMMENDATION: Because these are early days for undergraduate data science education, academic institutions should be prepared to evolve programs over time. They should create and maintain the flexibility and incentives to facilitate the sharing of courses, materials, and faculty among departments and programs.

Data science itself provides the tools to continuously evaluate and improve data science education. Evaluation should include assessment of student learning and assessment of how well a program is meeting the needs of the market it aims to serve. Evaluation can be used to shape a program at a given institution, showing what is working and where improvement is needed. It can also be used comparatively to detect approaches, classes, or curricula that may be of value to other campuses or contexts.

RECOMMENDATION: Academic institutions should ensure that programs are continuously evaluated and should work together to develop professional approaches to evaluation. This should include developing and sharing measurement and evaluation frameworks, data sets, and a culture of evolution guided by high-quality evaluation. Efforts should be made to establish relationships with sector-specific professional societies to help align education evaluation with market impacts.

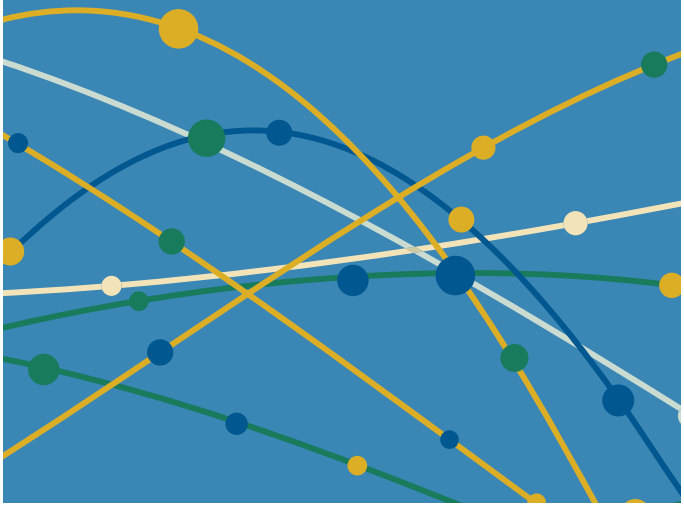
Much of the necessary data for evaluation could come from institutions' administrative records. These records, used in conjunction with other data sources such as economic information and survey data, could enable effective transformation and generalization of programs and might even inform a cohesive national approach to undergraduate data science education. In many fields, professional societies play a role in creating and nurturing community, in facilitating the sharing of resources and results, and in convening groups to set standards or determine best practices. Such capabilities are valuable to data science as well. However, it may be difficult for a single existing society to represent all the interests, disciplines, and professions included within the data science community. A structured collaboration of existing professional societies might work better, with potential development of subsocieties devoted to data science elements in preexisting societies.

RECOMMENDATION: Existing professional societies should coordinate to enable regular convening sessions on data science among their members. Peer review and discussion are essential to share ideas, best practices, and data.

Conferences, workshops, training sessions, and other networking opportunities would benefit the joint data science community. Collaborating societies could collect materials, create publication venues for the community, and convene discussions around critical topics such as curriculum, evaluation, and ensuring broad participation.

LEARN MORE

Download the full report and all report resources on the National Academies Press website at **nap.edu/25104**.



COMMITTEE ON ENVISIONING THE DATA SCIENCE DISCIPLINE: THE UNDERGRADUATE PERSPECTIVE: Laura Haas, University of Massachusetts Amherst, Co-Chair; Alfred O. Hero III, University of Michigan, Co-Chair; Ani Adhikari, University of California, Berkeley; David Culler, University of California, Berkeley; David Donoho, Stanford University; E. Thomas Ewing, Virginia Tech; Louis J. Gross, University of Tennessee, Knoxville; Nicholas J. Horton, Amherst College; Julia Lane, New York University; Andrew McCallum, University of Massachusetts Amherst; Richard McCullough, Harvard University; Rebecca Nugent, Carnegie Mellon University; Lee Rainie, Pew Research Center; Rob Rutenbar, University of Pittsburgh; Kristin Tolle, Microsoft Research; Talithia Williams, Harvey Mudd College; Andrew Zieffler, University of Minnesota, Minneapolis

This Consensus Study Report Highlights was prepared by the Computer Science and Telecommunications Board (CSTB), the Board on Mathematical Sciences and Analytics (BMSA), the Committee on Applied and Theoretical Statistics (CATS), and the Board on Science Education (BOSE) based on the report *Data Science for Undergraduates: Opportunities and Options* (2018). The study was sponsored by the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of the sponsors. Download the report at **nap.edu** and learn more about the study at **nas.edu/EnvisioningDS**.

STAFF: Michelle Schwalbe, Study Director and BMSA Director; Jon Eisenberg, CSTB Director; Ben Wender, CATS Director; Amy Stephens, BOSE Program Officer; Linda Casola, BMSA Associate Program Officer and Editor; Renee Hawkins, CSTB Financial Manager; Janki Patel, CSTB Senior Program Assistant

Division on Engineering and Physical Sciences

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.
www.national-academies.org