Establishing ABET Accreditation Criteria for Data Science

Jean R. S. Blair
United States Military Academy
West Point, New York, USA
jean.blair@westpoint.edu

Lawrence Jones
ABET, Inc.
Baltimore, Maryland, USA
lawrence.g.jones@ieee.org

Paul Leidig
Grand Valley State University
Allendale, Michigan, USA
leidigp@gvsu.edu

Scott Murray
HCA Healthcare, Inc.
Kennesaw, Georgia, USA
rsmurray1@gmail.com

Rajendra K. Raj
Rochester institute of Technology
Rochester, New York, USA
rkr@cs.rit.edu

Carol J. Romanowski
Rochester institute of Technology
Rochester, New York, USA
cjr@cs.rit.edu

ABSTRACT

Prompted by the skyrocketing demand for data scientists, progress made by the ACM Data Science Task Force on defining data science competencies, and inquiries about data science accreditation, ABET is in the process of developing accreditation criteria for undergraduate data science programs. The effort is led by members of a joint data science criteria subcommittee appointed by ABET's Computing Accreditation Commission (CAC) and CSAB (the lead society for computing accreditation). Establishing data science accreditation criteria is a notable milestone in the maturing data science discipline, indicating the presence of an accepted body of knowledge, standards of practice, and ethical codes for practitioners.

This position paper motivates the effort and discusses prior work towards defining data science education requirements. It describes the ongoing process for creating and obtaining approval of the accreditation criteria, and how feedback was and will be solicited from the computing and statistical communities. The current draft data science criteria, which was approved in July 2020 by the relevant ABET bodies for a year of public review and comment, is presented. These criteria emphasize the three pillars of data science: computing foundations, mathematical/statistical foundations, and experience in at least one data application domain. This report thus serves both to inform and to stimulate the academic discussion needed to finalize appropriate data science accreditation by ABET.

CCS CONCEPTS

• **Social and professional topics → Accreditation**; • **Mathematics of computing → Probability and statistics**; • **Information systems → Data management systems**.

KEYWORDS

Data science programs; data analytics programs; computing education; program accreditation

ACM Reference Format:

Jean R. S. Blair, Lawrence Jones, Paul Leidig, Scott Murray, Rajendra K. Raj, and Carol J. Romanowski. 2021. Establishing ABET Accreditation Criteria for Data Science. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21), March 13–20, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3408877.3432445

SIGCSE '21, March 13–20, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8062-1/21/03…$15.00
https://doi.org/10.1145/3408877.3432445

1 INTRODUCTION

In the past few years the demand for data scientists has skyrocketed. According to LinkedIn's Emerging Jobs Report [12] the number of data science jobs in 2020 will increase 37% in the United States alone; the European demand is similar [10]. Data science as a field has evolved since its beginnings over a half century ago when it was called the synthesis and analysis of data sets [19]; it now needs to deal with staggering quantities of data. More recently, there have been attempts to define data science as a discipline that is distinct from other computing programs, such as computer science or information systems, by its emphasis on the need for three distinct skill-areas (sometimes called pillars): computing, mathematics/statistics, and domain expertise [7].

Employers and data science professionals are interested in hiring competent data science graduates who have experienced this "three pillar" education; they also want to ensure that their hires adhere to standards such as the ACM Code of Ethics and Professional Conduct [5] when they handle data. The rapid proliferation of data science degrees makes it difficult to discern whether a program adequately prepares its graduates in this way.

Accreditation helps to address all of these issues through a process of external and impartial evaluation of degree programs. As professions mature, they create an accepted body of knowledge, standards of practice, and ethical codes for practitioners. Program accrediting bodies, such as ABET, verify that program graduates have received appropriate education in the discipline. Employers and students can thus be confident that graduates of an accredited program meet the educational requirements of the discipline. Due in part to the growth of data science as a discipline and a profession, concern about educating the data science work force has intensified. As a result, ABET was contacted both by institutions interested in accreditation for their data science programs and by employers wondering whether ABET was accrediting data science programs.

Consequently, ABET's Computing Accreditation Commission (CAC) and CSAB (the lead society jointly established by ACM and the IEEE Computer Society for all computing accreditation activities) formed a joint study group in 2018 to investigate the potential for accrediting bachelor-level data science programs. The initial step was to gauge industry demand for data science graduates while also researching the number of existing relevant undergraduate programs. The employer demand portion of the business case was easy to make. In the past few years, the demand for data scientists has skyrocketed, as noted above and by the LinkedIn's Emerging Jobs Report [12].

Although data science is inherently interdisciplinary, the study group found that a majority of the job listings emphasized computing skills [10, 16, 18]. Moreover, the study group's web-based investigation of 570 US institutions offering bachelor, master, minors, and certificate programs in data science showed that more than half had a computing emphasis. The second category of data science programs were roughly split between statistics/mathematics and business programs. Based on this study, ABET and CSAB decided to proceed by initially focusing on computing-oriented undergraduate data science programs that were potentially accreditable by the CAC. The study group is continuing to investigate the second category that includes both mathematical/statistical and business programs in the data science space.

This position paper details ABET's response to these developments and the process that is underway to establish data science program criteria. Section 2 examines the background in data science that helped shape the discussion of the data science discipline and the possible educational goals for data science programs. The ABET criteria development process is outlined, especially community engagement, in Section 3. The different elements of the draft version of the data science criteria, which is available for public review and feedback, are discussed in Section 4. We conclude with an overview of the next steps needed to finalize the data science accreditation criteria.

2 BACKGROUND AND RELATED WORK

Over 50 years ago, Tukey visualized the broad field of the synthesis and analysis of large data sets [19], which has evolved radically to the current world of data science (DS). In the early days, the focus was on mathematical and statistical analysis techniques, but exponential growth in data volumes in the modern world [17] has also required today's data scientist to be well-prepared in computing.

Building on Tukey's seminal work, several U.S. organizations have recently worked to speed up the process of developing a data science discipline and profession. The National Science Foundation (NSF) organized a workshop to gather diverse perspectives [6]. The Park City Math Institute developed interdisciplinary curriculum guidelines for undergraduate programs in data science [8]. The National Academies of Sciences, Engineering, and Medicine subsequently released a report that examined undergraduate data science education, attempting to define the discipline in a flexible way [13]. In 2018 a Data Science Leadership Summit [20] was held at the Data Science Institute in New York City to form an academic community for data science, share best practices, and educate the next generation of data scientists. The ACM Data Science Task Force was formed to explore the *broad, interdisciplinary conversation* on data science, and provide an articulation of the role of *computing discipline-specific contributions* to this emerging field.

Data science is an inherently interdisciplinary field that brings together domain data, computer science, and the statistical tools for interrogating the data and extracting useful information. Three essential elements of the data science environment are the **domain** that provides the data; **statistics** for analysis, modeling, and inference; and **computing** for data access, management, protection, as well as effective processing in modern computer architectures. Data science, as an interdisciplinary discipline, requires these three components to be integrated effectively to produce meaningful results.

The ACM Data Science Task Force is producing a report organized around competencies in eleven knowledge areas that are key computing contributions to data science [3]. A full data science curriculum needs to augment these computing knowledge areas with competencies in discrete structures, probability theory, statistics, and linear algebra, among others. A complete curriculum would also include at least one domain context for application of data science concepts and methods. The goal is to foster future collaboration to produce recommended guidelines for a complete data science curriculum including all of these competencies [3].

In Europe, the European EDISON project helped to speed up the establishment of data science as a profession [9] in 2015–17, especially through a first attempt to list data science degree programs, whose titles showed a great deal of disparity. Programs bore expected names such as data science, data analytics, data engineering and business analytics, but also included traditional names such as software engineering, computer information systems, and computer science.

The data world also benefited from Jim Gray's notion of data-intensive science as the fourth paradigm for 21st century science, following the first three: empiricism, theory, and computation [11]. Data science programs have also emerged in countries such as Australia, China, India and Singapore.

Figure 1 reflects the relative frequency in Google search trends under "Jobs and Education" for data science, big data, data analytics and big data analytics. The numbers on the y-axis represent search interest relative to the highest point on the chart for the given region and time. As Google Trends reports, for every 99 searches for *data science*, there were 24 for *data analytics* and only six for *data engineering* in August 2020; we concluded that Data Science has become the accepted term for this discipline.

The data thus show that the world of education has largely converged on the term *data science* for this interdisciplinary field. Data science extracts meaningful information from different kinds of data, throughout the entire data lifecycle, which includes acquiring, integrating, cleaning, analyzing and using data. In contrast, *data engineering* typically focuses on data infrastructure [15]; also, as shown in Figure 1, it has not reached the critical mass needed to be recognized as a separate discipline.

3 CRITERIA DEVELOPMENT PROCESS

In November of 2019, the ABET Board of Directors named CSAB as the interim lead society and the Computing Accreditation Commission (CAC) as the interim commission for data science programs, which prompted the Data Science Task Force to form a joint CSAB/CAC Data Science Criteria subcommittee. Members of the subcommittee were drawn from a cross section of industry and academia and included representation from programs that potentially could seek accreditation in the data science discipline.

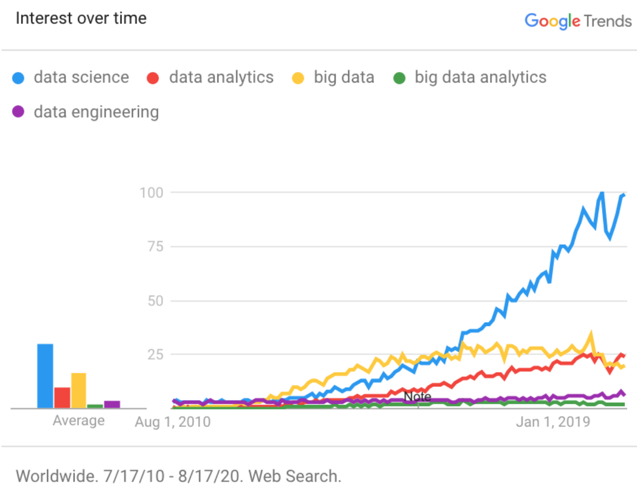

Figure 1: Search term popularity under Jobs & Education (Note: 100 means peak popularity for the term; 50 means half as popular; and 0 means insufficient data.)

The subcommittee was charged with establishing a robust, yet flexible set of criteria that supports program customization based on constituent needs. The charge further required that the criteria be resilient—likely to remain appropriate as a program criteria for years to come—and that it be designed so that assessment of student outcome(s) would not place undue burden on programs. While the focus was on developing criteria for computing programs, the committee was further asked to think about how other commissions might be able to use the same or similar criteria to accredit programs whose emphasis more closely associates with the other commission. For example, Data Science programs with heavy emphasis on mathematics or statistics might more appropriately be accredited by the Applied and Natural Sciences Accreditation Commission (ANSAC).

The Data Science Criteria Subcommittee goals included:

(1) establishing a set of criteria that would meet ABET approval,
(2) engaging repeatedly with the data science educational and professional community to gather feedback and suggestions,
(3) incorporating the relevant feedback into subsequent revised draft criteria, and
(4) working towards a potential cross-commission approach to criteria development.

CAC bachelor-level program criteria supplement the General Criteria [1]. The program criteria add an additional student outcome (Criterion 3) to the General Criteria's required student outcomes 1 through 5. Additionally, the Curriculum Criterion (Criterion 5) program-specific curricular requirements add to the General Criteria curriculum requirements. Therefore, certain topical areas, such as the "local and global impacts of computing solutions on individuals, organizations, and society," which are included within the General Criteria [1], are not repeated in the program criteria.

The subcommittee divided into two sub-groups, one focusing on the data science student outcome while the other group focused on curricular requirements. In spite of the fact that the discipline is rapidly evolving, the subcommittee was able to identify and use the following sources, which are converging toward a common view of the discipline:

1. EDISON Data Science Framework [9]
2. The National Academies vision of the Data Science Discipline [13]
3. Computing Competencies for Undergraduate Data Science Curricula – Draft 2 [3]
4. Park City Mathematics Institute Report [8]
5. An Empirical Approach to Data Science and Engineering Education [15]
6. An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation Experience [4]
7. NIST Big Data Interoperability Framework: Volume 1 [14]

The student outcome defined by the subcommittee succinctly articulates the unique (relative to other computing disciplines) skills that an undergraduate data scientist should possess upon graduation. The other major task was to capture the specific curricular topics expressly relating to required knowledge and skills areas for data science graduates.

During the criteria development process, the subcommittee collected a long list of suggested required knowledge areas but ultimately settled on the broad areas of the data science life-cycle, computing, mathematics, statistics, a program-selected domain area (or areas), and a major project requiring integration and application. This broad categorization helped meet the requirement to establish flexible criteria. The subcommittee further wanted to be sure to highlight the overarching nature of the computing, mathematics, and statistical topics in conjunction with data privacy, governance, and stewardship knowledge areas.

Throughout the process, the subcommittee sought formal and informal feedback on various drafts from a wide variety of constituents including academia and industry representatives along with the executive bodies of CSAB and ABET. In an effort to generate feedback from an even wider audience, several members of the subcommittee presented in-person at a pre-symposium workshop at SIGCSE 2020. In addition, the entire subcommittee conducted a virtual working session at the ABET Symposium in March 2020, with over 40 participants representing diverse computing and statistical departments offering data science programs. Lastly, in May 2020, the subcommittee offered two information and feedback sessions to the CAC membership.

All these sessions introduced the criteria development process, presented the then-current draft of the data science criteria, sought direct feedback from participants via a standard set of questions, and welcomed open-ended feedback from participants via email. The comments and feedback provided invaluable ideas for enhancement, including improved word selection, removing ambiguous words, reordering curriculum items for better flow and clarity, and streamlining the list of curriculum topics.

Currently, the data science criteria are open for public review and feedback until June 2021. All interested parties are encouraged to provide feedback on the criteria through the ABET website; meanwhile, the subcommittee will continue to seek feedback from a wide range of audiences and will submit the criteria for final approval by CSAB and ABET.

Table 1: Required CAC Student Outcomes (applicable to all Data Science programs) [1, 2]

	Graduates of the program will have an ability to:
	[The following five outcomes applicable to all computing programs]
1.	Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.
2.	Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
3.	Communicate effectively in a variety of professional contexts.
4	Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.
5.	Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline.
	[The following outcome is applicable only to Data Science programs]
6.	Apply theory, techniques, and tools throughout the data analysis lifecycle and employ the resulting knowledge to satisfy stakeholders' needs.

4 DATA SCIENCE CRITERIA

Since the Data Science Program Criteria build on the CAC General Criteria, a Data Science program must meet both the Computing Accreditation Commission (CAC) General Criteria and the Data Science Program Criteria.

The General Criteria covers eight areas: Students, Program Educational Objectives, Student Outcomes, Continuous Improvement, Curriculum, Faculty, Facilities, and Institutional Support. Only the Student Outcomes, Curriculum, and Faculty Criteria distinguish the computing General Criteria from the other commissions' General Criteria, and the Faculty Criterion is not significantly different.

4.1 Student Outcomes

The CAC Student Outcomes Criterion requires that a program have documented and publicly stated student outcomes that include the five outcomes required by the General Criteria for all computing programs, and one additional outcome required only by the Data Science Program Criteria. The program is free to have additional outcomes as well. Table 1 lists these six required student outcomes.

This criterion states that a data science program must:

- **adopt the six student outcomes verbatim**, with two allowed changes. In Outcomes 2 and 5 the phrase "the program's discipline" may be replaced by "Data Science." Additional program-defined outcomes are also possible.
- **publish their student outcomes** so that it is clear to the public that these are their student outcomes. It is appropriate to publish the student outcomes in all outlets that have a formal description of the program. Typically this would include at least the program's website and the program's description in the institutional catalog.
- **have a continual improvement process based on their student outcomes.** The process must be appropriate, meaning that it includes gathering relevant assessment data and using that data to effectively evaluate the extent to which each outcome is achieved by the program's collective student population. The process must be documented, i.e., written down, so that any evolving group of individuals responsible for it clearly understand it.
- **systematically utilize the results of the student outcome evaluations** as input for the program's continuous improvement actions. This includes documenting (recording) the decisions and actions taken during the continual improvement process.
- **ensure the curriculum is designed so that each student outcome can be attained.** This task is often accomplished by mapping the curriculum to the student outcomes.

4.2 Curriculum

The Curriculum Criterion for the General Criteria (GC) and Data Science Program (DS) Criteria are quoted in Tables 2 and 3 respectively. Note that the 45 semester credit hours of data science course work is meant to be a super-set of the 30 semester credit hours of up-to-date computing topics. Thus, together the GC and DS criteria require 45 semester credit hours (or equivalent) that include items 1-3 in Table 2 and items 1-5 in Table 3.

Table 2: CAC General Criteria Curriculum Requirements (applicable to all computing programs) [1]

	CAC General Criteria—Curriculum:
	The program's requirements must be consistent with its program educational objectives and designed in such a way that each of the student outcomes can be attained. The curriculum must combine technical, professional, and general education components to prepare students for a career, further study, and lifelong professional development in the computing discipline associated with the program.
	The curriculum requirements specify topics, but do not prescribe specific courses. The program must include mathematics appropriate to the discipline and at least 30 semester credit hours (or equivalent) of up-to-date coverage of fundamental and advanced computing topics that provide both breadth and depth.
	The computing topics must include:
1.	Techniques, skills, and tools necessary for computing practice.
2.	Principles and practices for secure computing.
3.	Local and global impacts of computing solutions on individuals, organizations, and society.

The three curricular items in the General Criteria are meant to be applied in the context of a computing discipline, which in this case is data science. Item 1 emphasizes relevant hands-on practice using up-to-date techniques, skills, and tools. Examples include use of integrated development environments (IDEs), debugging skills and tools, version-control tools, data analysis and statistical modeling tools, machine learning tools, and data visualization tools. Item 2 could be accomplished with either a required secure computing course that includes both security principles and security practice or by weaving security practices and principles throughout multiple required courses. Item 3 requires that the students have learning experiences in which they analyze the impact of data science solutions in a range of contexts. These might address, for example, different cultural, economic, ethical and legal impacts; impacts on

Table 3: Data Science Curriculum Requirements (applicable only to all data science programs) [2]

Data Science Program Criteria—Curriculum:
The curriculum requirements specify topics, but do not prescribe specific courses.

These requirements are at least 45 semester credit hours (or equivalent) of data science course work that must cover:

1. Fundamental data analysis life-cycle topics:
   a) Data acquisition
   b) Data management
   c) Data preparation and integration
   d) Data analysis
   e) Model development and deployment
   f) Visualization
2. Concepts that span and are applied to the data analysis life-cycle:
   a) Data privacy, governance, and stewardship
   b) Statistics and mathematics
   c) Computing, including substantial coverage of data structures, algorithms, and at least one programming language
3. Advanced data science coursework that provides depth
4. Coverage of at least one application domain area to provide a context for data science activities
5. A major project that 1) incorporates an application domain area and 2) requires integration and application of knowledge and skills acquired in earlier course work

diversity and inclusion, privacy, and accessibility; or impacts when solutions are deployed to different regions of the world.

The five curricular items in the Data Science Criteria focus on knowledge and skills related to the data analysis life-cycle and being exposed to a context for applying those skills. Item 1 explicitly lists six data analysis life-cycle topic areas. Item 2 emphasizes the inter-disciplinary nature of the discipline, including statistics, mathematics, computing, and professional responsibilities. Of note is the fact that the application domain in Item 4 does not have to be the same as the major project's application domain in Item 5.

5 NEXT STEPS

The Data Science Program Criteria version described in this report is the first-reading version, meaning that the criteria are now available for public review and comment [2]. Feedback will be considered and appropriately incorporated into what is called a second reading prior to ABET's summer 2021 commission meetings. If the second reading program criteria is the approved through the ensuing ABET processes, the criteria will be available for accreditation of data science programs by ABET's Computing Accreditation Commission starting in the academic year 2022-2023 accreditation cycle.

Additionally, an effort is underway to establish a "harmonized" set of Data Science Program Criteria under the Applied and Natural Sciences Accreditation Commission (ANSAC). This work includes researching potential programs that could seek accreditation from ANSAC, developing harmonized data science criteria for use by ANSAC, and establishing procedures for determining the appropriate accreditation commission for data science programs seeking initial accreditation.

Investigation is ongoing on the nature of data science programs that are offered by mathematics/statistics departments or by business schools. However, the results of the study is beyond the scope of this position paper, which has been to discuss the process used and status of ABET data science accreditation criteria for undergraduate programs that emphasize computing skills.

ACKNOWLEDGMENTS

Our thanks are due to Lillian (Boots) Cassel, Hridesh Rajan, Barbara Price, and Heikki Topi, who served on the Data Science Criteria Subcommittee and helped to develop the Data Science Criteria proposal drafts. We also thank the many folks who attended the SIGCSE 2020 pre-symposium Data Science workshop and the ABET 2020 Symposium Data Science Working Session and provided input and feedback into earlier drafts of the criteria proposal as it evolved over the past three years. Rajendra Raj acknowledges support provided by the US National Science Foundation under Awards 1433736, 1922169, and 2021287.

REFERENCES

[1] ABET, Inc. 2020. Criteria for Accrediting Computing Programs. Effective for Review During the 2020-21 Accreditation Cycle. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-computing-programs-2020-2021/.

[2] ABET, Inc. 2020. Program Criteria for Data Science and Similarly Named Programs, 1st Reading. https://abet.co1.qualtrics.com/jfe/form/SV_02pWHHMsKN46j9b

[3] ACM Data Science Task Force. 2019. Computing Competencies for Undergraduate Data Science Curricula. https://dstf.acm.org/

[4] Paul Anderson, James Bowring, Renée McCauley, George Pothering, and Christopher Starr. 2014. An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation Experience. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education* (Atlanta, Georgia, USA) *(SIGCSE '14)*. Association for Computing Machinery, New York, NY, USA, 145–150. https://doi.org/10.1145/2538862.2538936

[5] Association for Computing Machinery. 2018. ACM Code of Ethics and Professional Conduct. ACM, New York. Webpage: https://www.acm.org/code-of-ethics.

[6] Lillian Cassel and Heikki Topi. 2015. Strengthening Data Science Education Through Collaboration. Workshop on Data Science Education, October 2015.

[7] Andrea Danyluk, Paul Leidig, Lillian Cassel, and Christian Servin. 2019. ACM Task Force on Data Science: Draft Report and Opportunity for Feedback. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. ACM, New York, NY, USA, 2.

[8] Richard D. De Veaux, Mahesh Agarwal, Maia Averett, Benjamin S. Baumer, Andrew Bray, Thomas C. Bressoud, Lance Bryant, Lei Z. Cheng, Amanda Francis, Robert Gould, Albert Y. Kim, Matt Kretchmar, Qin Lu, Ann Moskol, Deborah Nolan, Roberto Pelayo, Sean Raleigh, Ricky J. Sethi, Mutiara Sondjaja, Neelesh Tiruviluamala, Paul X. Uhlig, Talitha M. Washington, Curtis L. Wesley, David White, and Ping Ye. 2017. Curriculum Guidelines for Undergraduate Programs in Data Science. *Annual Review of Statistics and Its Application* 4, 1 (2017), 15–30. https://doi.org/10.1146/annurev-statistics-060116-053930

[9] EDISON Poject. 2017. Data Science Competence Framework (CF-DS). https://edison-project.eu/data-science-competence-framework-cf-ds/

[10] Harnham Search and Selection Ltd. 2020. The Harnham Data & Analytics Salary Guide: European, UK and US. https://www.harnham.com/2020-data-and-analytics-salary-survey.

[11] Tony Hey, Stewart Tansley, and Kristin M. Tolle (Eds.). 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft Research, Redmond, Washington. http://research.microsoft.com/en-us/collaboration/fourthparadigm/

[12] LinkedIn. 2019. Emerging Jobs Report. https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/emerging-jobs-report/Emerging_Jobs_Report_U.S._FINAL.pdf.

[13] National Academies of Sciences, Engineering, and Medicine. 2018. *Data Science for Undergraduates: Opportunities and Options.* The National Academies Press, Washington, DC. https://doi.org/10.17226/25104

[14] NIST Big Data Public Working Group,Definitions and Taxonomies Subgroup. 2019. *NIST Big Data Interoperability Framework: volume 1, definitions version 3.* Technical Report. National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.SP.1500-1r2

[15] Rajendra Raj, Allen Parrish, John Impagliazzo, Carol Romanowski, Casey Bennett, Sherif Aly, Karen Davis, Teresa Pereira, Andrew McGettrick, and Lovisa Sundin. 2019. An Empirical Approach to Understanding Data Science and Engineering Education. In *roceedings of the Working Group Reports on Innovation and Technology in Computer Science Education (ITiCSE-WGR '19)*. Association for Computing Machinery, Aberdeen, UK. https://doi.org/10.1145/3344429.3372503

[16] Steve Ranger. 2019. Python vs R and Biggest Salaries: Top Data Science Job Trends. https://www.zdnet.com/article/python-vs-r-and-biggest-salaries-top-data-science-job-trends/.

[17] David Reinsel, John Gantz, and John Rydning. 2018. Data Age 2025: The Digitization of the World from Edge to Core. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf.

[18] Arwa Sutarwala. 2019. Three Trends in Data Science Jobs you should know. https://dataconomy.com/2019/07/three-trends-in-data-science-you-should-know/.

[19] John W. Tukey. 1962. The Future of Data Analysis. *The Annals of Mathematical Statistics* 33, 1 (03 1962), 1–67. https://doi.org/10.1214/aoms/1177704711

[20] Jeannette M. Wing, Vandana P. Janeja, Tyler Kloefkorn, and Lucy C. Erickson. 2018. *Data Science Leadership Summit: Summary Report.* Technical Report. Data Science Institute, USA. https://dl.acm.org/citation.cfm?id=3293458