Executive Summary

The Presidential Task Force on Applied Data Science is investigating Seattle University’s involvement in a relatively new and rapidly evolving technological and collaborative field. While there is a lack of congruent language across the many academic fields that intersect with and impact applied data science, the disciplines include data analysis, data visualization, data project management, digital marketing, applied statistics, chemometrics, healthcare informatics, database architecture, algorithms, software, machine learning, deep learning, and artificial intelligence (etc.). These disciplines all contribute to the intellectual capacity of the new data-smart student Seattle University hopes to serve.

We believe that training in the computational tools of today represents a fundamental skill for students in all fields. While technology and the scope of problems we can investigate are certainly skyrocketing, the central capability of any scientist or decision maker is still the scientific method: the ability to form a testable hypothesis that connects us to an in-depth understanding of our world. We stand at an exciting time where technology and data are accessible to all. However, learning how to glean useful information from data still requires careful training and education.

The successful implementation of data-focused education at Seattle University faces six core challenges as identified by the Task Force.

1. How do we integrate data literacy into undergraduate education?

2. Are there more opportunities for graduate level education?

3. With its highly interdisciplinary nature, are there ways to provide students with this training at Seattle University?

4. What would a University Center of Applied Data Science look like?

5. What infrastructure is needed?

6. What is the current supply and demand for data-focused education?

The Task Force has explored what it would take to lead Seattle University to the forefront of undergraduate and graduate education in applied data science. Because of the scope and relative newness of this emerging field, idealism is leading the way with much collegial optimism and excitement; however, there is also a recognition of practical constraints.

Educational options for Seattle University focus on the need for best practices and the appropriate design and timing of data literacy education. Also, delivering engaging content with real world data is a must (multiple employers highlighted this as a primary need). There are possible ways to integrate an undergraduate minor that could add some interdisciplinary data analytics knowledge to a student’s degree focus. However, we acknowledge that integration into the existing University CORE curriculum and development of graduate-level interdisciplinary certificate programs, and/or a full-fledged graduate degree appear to be the educational paths that are most promising.
Practically, a physical or virtual collection of resources and experts are required here at Seattle University, the specifics of which will likely be limited by available funding. Some staffing for resource management, training, and facilitating open-collaboration across departments is critical to success, as well as the ability to coordinate exciting events such as a data science hackathon. The campus needs a coordinated effort to bring disciplines together and connect students to projects and resources. Currently, faculty and students are making partial use of cloud resources as best they can, but there is much needed support outside of what is currently in place.

The recommendations of the Task Force are as follows:

**University Core Curriculum and Data Literacy Education**

**Recommendation #1:** Leverage the upper end of the University Core by focusing on the 3000 level, developing a suite of “3200” level courses, and/or by focusing on the Global Challenges courses.

**Recommendation #2:** Expand the lower end of the University Core by taking advantage of existing processes (mechanics) for Core expansion and development.

**Recommendation #3:** In recognition of the fact that data literacy education can also be developed, introduced and sustained by the provision of courses and/or tracks outside of the University Core, exploration of the provision of courses and/or tracks through the “Center” or through New and Continuing Studies is recommended.

**Recommendation #4:** Exploration of the potential of data literacy or data science boot camps or workshops is also recommended. Providing immersive, hands-on, short-duration experience, boot camps could be sponsored by the “Center,” New and Continuing Studies or other schools and/or colleges within Academic Affairs.

**Joint Academic Programs in Applied Data Science**

**Recommendation #1:** The Subcommittee recommends that Seattle University develop an interdisciplinary undergraduate minor in applied data science. This minor should include foundations in computer programming, databases, statistics, and data visualization.

**Recommendation #2:** The Subcommittee recommends exploring the possibility of a graduate certificate or master’s program in data science. This should be done in consultation with whatever college(s) / department(s) would host such a program (such as Mathematics or Computer Science in the College of Science and Engineering).

**Recommendation #3:** The Subcommittee recommends that any development of graduate programs in data science consider the feasibility and potential demand for dual degree programs with data science and existing graduate programs such as criminal justice or business analytics.
Graduate Curriculum

Recommendation #1: Developing of an interdisciplinary, comprehensive, stand-alone Introduction to Data Science course for Seattle University graduate students who have no programming background.

Recommendation #2: Develop a “Programming Boot Camp” aimed at students who have no prior programming experience, but who will benefit in their studies and work from developing a knowledge of a programming language such as Python. The Computer Science department currently offers a programming boot camp (https://www.seattleu.edu/scieng/computer-science/certificates/cs-fundamentals/programming-boot-camp/ (accessed 25 September 2018)) aimed at students intending to do intensive software development. There is need for a similar program for students who will not be going into computer science, but who do need a working familiarity with a high-level programming language. The structure and details of the new boot camp should be determined by a set of representatives of graduate programs that would subscribe to the boot camp.

Recommendation #3: Ensuring students in all Seattle University graduate programs have the necessary campus resources for assisting them with large dataset collection, analysis, and dissemination.

Establishing a Seattle University Applied Data Science Center

Recommendation #1: Commit to establishing a Seattle University Applied Data Science Center. Provide resources to the center from internal sources as well as from external sources such as industry, foundations, and individual donors.

Recommendation #2: Establish the Center as a hub for all data science activities. This would include:
- Identify the initial collaboration platform(s) to be promoted by the Center for use by faculty, students, and staff in staying abreast of currently available capabilities.
- Reach out to faculty with existing teaching and research obligations to assure that needs are properly assessed and tracked.
- Provide general guidance to the Seattle University community regarding the use of available Applied Data Science vendor/product/service capabilities. Share how input for changes and improvements will occur.
- Establish Applied Data Science governance relationships and processes. Governance would include: decision-making regarding desired products, services and tools for Applied Data Science; formal requests for prioritized infrastructure investments; policy recommendations; intellectual property guidance; and other relevant decisions. Governance partners would include the Director of the Applied Data Science Center, the Faculty Technology Committee, the Chief Information Officer, and others.
**Recommendation #3:** Ensure adequate infrastructure and resources for applied data science initiatives. This would include:

- Establish an annual budget for Applied Data Science infrastructure investments and identify an appropriate source of funding.
- Hire a technical resource within IT Services (1 FTE) to focus on cloud infrastructure implementation for Seattle University. This resource would implement enabling technologies (e.g., network, identity, security) and partner with the practitioners/coaches in the Applied Data Science Center and the various academic programs.
- Investigate and assess the need for an Applied Data Science Librarian.
- Create and maintain a multi-year rolling roadmap of planned Applied Data Science infrastructure investments (includes back-office technologies, end-user tools/product/services purchases) and deliver these improvements per the master schedule. Examine if a phased approach might be a rational method for implementing infrastructure investments.

Overall, there is growing demand for experts in data from a wide variety of backgrounds. Excitingly, the surrounding industry in Seattle is one of the intellectual centers of this emerging field and many opportunities exist to develop an exceptional and uniquely customized program here at Seattle University. Our students stand in front of a tremendous amount of data, no matter what field they choose to enter. Seattle University’s consistent commitment to serve the student requires igniting youthful open-collaboration to face this onslaught of data with mathematical precision and social clarity.

The city/region is rich with opportunities for immediate and long-term investment and engagement opportunities that align with and advance our vision. For our initiative to be successful, the university must explore opportunities to partner with industry, foundations, and individual donors who will support applied data science initiatives, sometimes for mutual benefit, and sometimes to support a compelling educational initiative. The concept of a cross campus center in applied data science presents a distinctive opportunity to garner support for our work.
Task Force Charge, Members, and Process

In November of 2017, Fr. Sundborg convened a Presidential Task Force charged with designing a strategy for how the curriculum and academic programs at Seattle University can best respond to technological advances in areas related to Big Data / Data Science—which broadly understood includes data analysis, data visualization, data project management and social media management.

The Applied Data Science Task Force consists of the following members:

- Chris Dale, Chief Quality Officer, Pulmonary & Critical Care Medicine, Swedish Health Services
- Griffin Deebach, Director of Analytics, Triomphe Global Marketing
- Gareth Green, Associate Professor of Economics, Albers School of Business and Economics
- Matt Hickman, Associate Professor of Criminal Justice, College of Arts and Sciences
- Helen Jiang, Senior Analyst and Technical Fellow, Boeing Commercial Airplanes
- Lin Li, Assistant Professor of Computer Science and Software Engineering, College of Science and Engineering
- Luke Marney, Instrumentation Manager, Department of Chemistry, College of Science and Engineering
- Patrick Murphy, Associate Professor of Nursing, College of Nursing
- Joe Phillips, Dean, Albers School of Business and Economics (Chair)
- Chuck Porter, Retired CIO, Seattle University
- Shawn Rider, Director of Web, Applications, and Technology Studies, School for New and Continuing Studies
- McLean Sloughter, Associate Professor of Mathematics, College of Science and Engineering
- Brendon Taga, Associate Dean, College of Education
- Steve Tapia, Distinguished Practitioner in Residence, School of Law
- Chris Van Liew, Chief Information Officer and VP for Information Technology
- Sarah Barbara Watstein, Dean, Lemieux Library and McGoldrick Learning Commons
Definitions, Context, and Outlook

Highly interdisciplinary, Data Science is a newly emerging academic field. Much as computer science and statistics each grew out of roots in mathematics to ultimately become influential fields in their own rights, data science is currently in the transitional process of becoming its own unique discipline, with roots in statistics and computer science.

As a newly emerging field, there is not yet a single unambiguous definition of what data science is. It is most commonly understood to refer to leveraging the combined tools of computer programming, databases, statistics, and data visualization to utilize data to answer real-world questions. Data scientists extract useful information and gain insight from huge data; the main objective of this work is to reveal the characteristics of natural, human, and social phenomena using data. Data scientists, as a broad category, could include business intelligence analysts, data engineers, and research scientists.

Data science – and data science competency, have developed in response to the ubiquity of data in modern life. Data has become a part of almost every career or field of study, and this has created a demand for expertise in working with and analyzing data in ways that are both methodologically sound and focused on practical application.

While there is growing demand for data scientists as a group of specialists in these methods, there is also a strong need for professionals of a wide variety of backgrounds to have some skills and training in data science. Social scientists, journalists, criminal investigators, lawyers, and others are making use of data science tools in their fields. Consequently, there is a growing need for academia to provide proper training for them in data science while they are students regardless of their interests, backgrounds and intended major.

Data science is also a rapidly evolving field. As our technology continues to develop, so too must the ways in which we engage with data. As we train our students to enter into this newly data-driven workforce, we must be equipping them to be able to flexibly adapt to new data analysis tools and methodologies they will encounter throughout their career, and we must ensure that our own programs within Seattle University are continually adapting in response to ever-changing industry needs.

Seattle, with its deep connections to the technology industry, has become one of the intellectual centers of this emerging field of data science. There are now many private boot camp programs offering to train technology workers and non-technical audiences alike in data science. At the same time that Seattle University was launching our data science intensive Masters of Science in Business Analytics program, The University of Washington was launching a more computer science focused data science program. Several other universities in the area are in the process of developing undergraduate and graduate programs in various aspects of data science. There is strong demand and interest for data science in the region, across a variety of focuses and areas of application.

Given the technology industry in Seattle, and our university’s existing successes with industry partnerships for student projects, there is an opportunity for our university to develop a program with strong industry ties that would not be available in other parts of the country.

In addition, with Seattle University’s smaller size, which often leads more easily to interdisciplinary work, we are well-positioned to embrace data science and develop an approach
to data science that is strongly interdisciplinary, that draws on the varied strengths of our faculty
who are already engaged with data science in statistics, business, computer science, law,
criminal justice, environmental science, and other fields. We have an opportunity to create a
unified field of study in data science amongst these varied faculty members which will benefit
our students across all fields of study.

In doing its work, the Task Force confronts several tensions. These tensions are depth vs.
breadth, focus vs. dissemination, and standardization vs. experimentation. These are not unique
tensions to Seattle University. Indeed, there have been many articles and anecdotal stories
about how colleges and universities are responding to the demand for data science. Higher
education institutions nationwide are asking, “Should we develop a data science program?” As
institutions explore how to respond to the demand for data science, they inevitably confront
similar tensions. Within these tensions lies the correct path forward.

First, depth vs. breadth. Deep offerings produce highly skilled graduates, with significant
practical experience in relevant tools, platforms, and/or methodologies, who are ready to be
employed as data scientists in tech savvy companies regionally and nationally. Such training
and course work will require significant prerequisites, including deep math and programming
language(s); these might become a barrier to entry for many students. On the other hand,
shallower offerings, such as what could be introduced into the university undergraduate core
curriculum, would provide students with foundational data-literacy skills and competencies, but
fall short in providing our students with technical depth and practical data science hands-on
experience. This is likely to be a barrier to hiring for actual data science roles regionally or
nationally. In the same way that a survey of science or intro to [fill in the discipline] rounds out
students, a survey of data science or intro to data science would round out students and be
positive, but might not position them to compete for professional opportunities or further
distinguish the university.

A second tension concerns focus vs. dissemination. Should we attempt to spread applied data
science throughout the schools’ and colleges’ programs and degrees, or focus it more highly in
a few offerings? Dissemination could move the entire university incrementally forward.
Dissemination also has the advantage of spreading risk and letting the market decide where
greater focus might pay dividends later on. Focus would allow resources to be concentrated into
a few bigger bets and allow for depth. Focus is probably a more efficient financial strategy, and
it would enable the university to build a reputation around distinctive programs which would be
highly visible externally. However, the risk is also more concentrated. Additionally, there is also
the problem of leaving some schools and colleges disenfranchised. It remains to be seen
whether focus will discourage innovation across the institution and lead to expensive, inefficient
and invisible shadow programs.

The final tension is standardization vs. experimentation. Do we want to encourage
experimentation throughout the University, facilitating rapid innovation in the small? Or, do we
want to create coherent standard tools, practices, and support which might facilitate innovation
in the large? We have many faculty interested in data science and they are adopting freemium
hardware/software/compute resources and experimenting around their own particular interests.
This has the potential to push into a new field quickly, but requires pruning failed experiments
and uneconomic strategies, and it requires knowing when to scale up those experiments that
worked. Standardization is efficient, minimizes financial or technical risks, is manageable and
administrable, projects power and progress, and seems more likely to move the larger needle.
Still, obtaining commitment to standards can be challenging and pushing faculty in any direction
has its own special risks and challenges.
Reviewing the Data Science Landscape

To assess the need for applied data science initiatives at Seattle University, the Task Force scanned the emerging needs of employers as well as how other institutions of higher education are responding to those needs. In particular, this subcommittee is seeking to answer two questions:

- What do businesses, nonprofits and other organizations who hire our students need for their earlier-in-career positions?
- What are those who train and otherwise develop the supply of talent for these positions doing in the market?

Summary of Market Place Analysis: Demand

To assess what the market requires for talent in the Applied Data Science realm, we used internet-based research regarding recruitment by organizations and the published research on this topic, as well as interviews with select local company leadership and facilitating intermediaries such as Seattle University’s Career Services unit. The analysis at this point is necessarily US-centric, as there has not been time to develop a fully international viewpoint on demand. Our analysis was also informed by the Business Higher Education Forum analysis of the need for data science talent in US business. It concludes the demand for data science jobs is increasing, and given the shortage of job seekers with meaningful data science skills and experience, salaries are increasing as well.

General Market Demand for Data Science and Analytics

We now measure the volume of newly created data in the quintillions of bytes each year, and there has been a rapid escalation in the number of open job positions in the marketplace that require data science-related skills, from data-driven decision makers to the data scientists and others who produce the data they require. The more technical data science-related annual job postings in the US alone number in the hundreds of thousands, with a high percentage of positions in professional, scientific and technical services, information services, manufacturing, health care, finance and insurance, and retail. A growing number of data analytics and data science positions are now found in state and local government, as well. Geographic hiring spans the major employment centers of the US.

Categories of “Data Scientists”

Analyzing the market demand for “data science”-related positions is complicated by the varying roles and career trajectories found within this broader category. Data Science positions have many dimensions, including: statistics, mathematics, big data/data engineering, machine learning and algorithms, business, software engineering, data visualization, spatial data, etc. Data science elements are also becoming an important component of many roles, including data-driven decision makers and functional analysts, but the primary focus of this specific analysis is limited to full time data science-like roles. Analyzing data science-related positions is complicated by the lack of standard titles used by organizations in recruitment efforts. However, by focusing on the entry point for the myriad data science roles, several primary entry-level trajectories come into focus:
• Data/BI Analyst: Enters the Applied Data Science workforce with SQL, extract, transform and load (ETL), data visualization, business, and at least some limited level of programming/scripting skills and competencies with technology infrastructure, including cloud provider services. Positions require at least a bachelor’s degree. Career path may grow into “data scientist” with additional training and experience.

• Data Engineer: Enters the Applied Data Science workforce with the above, but also skills in data architectures, data mining, data plumbing, operations research, probabilities, statistics, predictive analytics, business, and more. Additionally, data engineers are expected to have deeper technology infrastructure and programming skills than data analysts, and be strong in Python and/or R in particular. Advanced degrees preferred and often required. A trend in the market is to retitle this role as “data scientist” at the entry level in an attempt to aid recruitment.

• Research Scientist: Enters the Applied Data Science workforce especially strong in math, statistics, and algorithms, with a deep specialization such as distributed systems, machine learning, pattern recognition, artificial intelligence, or other. A given individual may have a selection of skills held by the analyst/engineer roles noted, above. Role often focuses on innovation and new product design, but responsibilities vary by organization, especially at “entry” level. Advanced degrees required.

In general, data analysts and data engineers are business-savvy with strong communication skills and will typically act as a “bridge” role with business leadership, helping them understand the implications of data analysis. Additionally, data engineers understand the needs of research scientists, and typically work to solve the data cleansing and data availability issues for research scientists, who spend most of their time focused on the math challenges. The research scientist position will tend to work primarily with other data science-related roles, but not exclusively.

As mentioned, the list above does not fully represent the nature with which myriad roles within organizations will require “data skills,” and how those requirements will increase over time. Just as skills in typing, then word processing, then spreadsheets, presentations, email and calendaring became part of the basic skills expected of nearly all employees, it seems likely that all employment disciplines will ultimately require data-related skills. We expect Seattle to be on the forefront of these changes. These demands will likely necessitate inclusion of data-related skills development at some level in all SU programs for our graduates to remain competitive.

Experience Required

Based on a broad review of positions posted in the marketplace, it is actually very difficult to find many truly “entry-level” data science-related positions. In other words, most positions in this category prefer, and many times require, 1-3 years of experience demonstrating the skills required for data/BI analyst, data engineer, and the research scientist positions. For example, many organizations find that their data is poorly structured and frequently has low data quality, integrity and veracity, and they want staff who can hit the ground running to resolve these issues, even in entry-level positions. In discussing this topic with local leaders, they emphasized the need for internships and meaningful projects that maximize the time doing hands-on work, especially with very large datasets. They also validated the value of an Applied Data Science Center for Seattle University that could partner with industry and potentially provide opportunities for consulting by the students to gain experience prior to graduation. These steps will help our graduates have the best chance of making a smooth transition from education to employment.
In a discussion with an Amazon leader, she noted, “Most students have the theory, but do not have sufficient tool experience. Students need practical experience in ETL, SQL, Hadoop/map-reduction software, large datasets (which Amazon can provide), and more.” She added, “The technical competency bar is VERY high for Amazon. An Amazon “entry level” person is more like a mid-level role anywhere else. This profile necessitates that the candidate has solid hands-on experience.” Along these lines, many Seattle University faculty have also affirmed their desire that students have the opportunity to work on real projects for local businesses in their coursework in order to be properly prepared with theory and experience for the organizations who hire them.

Further Research

It might be possible to collaborate with larger employers, especially regionally, to design data science-related programs and other student experiences as an explicit entry point into their data science roles. Further research regarding corporate interest in this approach or other collaborations is recommended.

Summary of Market Place Analysis: Supply

To assess what the market offers in the Applied Data Science realm, we used primarily Internet-based research methods to examine what Higher Ed offers and what private programs offer. We chose this method because students looking for training in Applied Data Science would follow this approach to find them and if programs were not revealed on the Internet, they would likely not attract students.

Graduate Programs

There are many well-respected graduate programs from well-respected universities that offer advanced degrees focused on Data Science. None are meant as “degree completion” or “certificate” programs. A majority offer both full-time, as well as part-time programs. All part time programs are completed, at least partially, on-line. Full time programs tend to be face-to-face on-campus programs. Most programs are housed within Engineering, Computer Science or Business.

Nearly all offer practical experience with Data Science: Experiential capstone courses; Internships with local businesses (some Fortune 500 companies); or both. Programs vary in their intention and rigor. One third are deep Computer Science leading to PhD or highly technical careers. Two thirds are more mercenary, aimed at applying Data Science in business careers. About half are oriented toward working professionals and half toward full-time residential students. Most of them offer dedicated cloud resources (e.g., Microsoft Azure, Amazon Web Services) or on-campus servers to provide a working computer environment for study.

Undergraduate Programs

It has been only recently that undergraduate degree programs in Data Science have become available. In general, institutions with existing Computer Science, Informatics, or Business Intelligence.Business Analytics have begun by adding specific Data Science classes in these programs (most institutions, today). As classes are added and become more robust, specific
concentrations for Data Science are branded and marketed within those programs, or the marketing for the degree is updated with a predominantly Data Science spin to it (e.g., UW, Santa Clara, Arizona State). Some of these programs are also available in a fully on-line format. Multiple institutions have now created full Data Science undergraduate degree programs, as well (e.g., WSU, San Francisco, UC Berkeley). Finally, a long list of institutions have an available minor in Data Science (sometimes alternatively titled “analytics” or “informatics”), which is typically an interdisciplinary minor offered to most majors in the institution to enable students to build skills to apply Data Science to their chosen discipline (e.g., UW, Stanford, MIT, Notre Dame, Oklahoma).

We expect a steady growth in the number of Data Science undergraduate degree programs available to prospective students at both state and independent institutions.

**Boot Camp and Certificate Programs**

By definition, these programs are not intended to confer degrees. Rather they promise to prepare students for lower level jobs in Data Science fields or for improved prospects at existing employers.

Specialized technology programs (or Boot Camps) such as Code Fellows, General Assembly, Coding Dojo, Tech Academy, Epicodus – are all available in the Pacific Northwest. Except for General Assembly, these tend to be short programs (6+ weeks), offered either in person or online, and centered around tech-savvy major cities. They tend toward software, web and mobile development and are not focused specifically on data science. They would tend to build programming prerequisites for data science study but not the math and statistics. They do not offer a “certificate” per se.

General Assembly *does offer a data science curriculum* that looks reasonably good. It includes Statistics lite, Data Modeling lite, and “Data Modeling in the real world.”

There are many certificate programs such as: Dell EMC Proven-Professional certification program; Certified Analytics Professional (CAP); SAS Academy for Data Science; Microsoft Certified Solutions Expert (MCSE); Cloudera Certified Associate (CCA); Cloudera Certified Professional: CCP Data Engineer; Data Science Certificate – Harvard Extension School. Some of these are testing programs – Students pay a fee, take the test, pass and receive the certificate. (Dell, CAP, Microsoft, Cloudera would fall into this category.) Some of these are training programs which produce a certificate at the end (Harvard ($15,000) and SAS ($20,000) fall into this category and charge for the training.)

Many universities offer non-degree certificates, often through their Continuing Education or Extension programs such as: UW (CE); Harvard (Extension); Cornell (Online Certificate); UC Berkeley (Extension); UC Irvine (CE); Ohio State (Online Certificate); and University of Massachusetts – Amherst (Certificate).

**Further Research**

It is recommended that the Task Force identify key graduate, undergraduate, certificate and other programs to monitor and leverage as examples for Seattle University’s consideration as we move forward.
Additional Questions

There are many additional questions that could be pursued related to this analysis. The following will be addressed in the paragraphs that follow:

1. Do we have demographic trend data that validates increased demand and as a consequence, increased compensation, for data science-related roles?
2. Beyond full-time roles, is data science exposure viewed by employers as a differentiator for other roles that are not full-time data science positions?

What is the Likely Trend for Hiring in Data Science?

The McKinsey Global Institute’s Report, Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation, speaks to how automation has changed the demand for data analysis-related roles: (p. 42)

In theory, at least, many information analysts might have been replaced by the rise of the Internet, which makes collecting data and information vastly more efficient, and computers that enable rapid and complex computations. After all, much of their work in the 1980s, before these technologies were widespread, has since been automated.

In fact, the opposite has happened. Computers and the Internet automated activities such as basic mathematics and information gathering, yet the number of information analysts soared nonetheless. As computers became more efficient, the cost of obtaining high-quality information dropped. Rather than decreasing demand for analysts, this stimulated the appetite for more insightful and low-cost analysis, and the number of analysts quintupled from around 400,000 in 1980 to about two million today (Exhibit 10).

The jobs of analysts have changed as well. With information and data more easily accessible, analysts can focus on making sense of—and sharing—information rather than gathering it.

The specific job titles that include a need for data science-related skills include Credit Analyst, Budget Analyst, Financial Analyst, Statisticians, Operations Research Analysts, Economists, Market Researchers, Survey Researchers, and Management Analysts. Additional data science roles are being added to each industry vertical rapidly, as the value of monetizing additional data insights are identified.

The report, Investing in America’s Data Science Talent: The Case for Action, created by Price Waterhouse Coopers, and published in conjunction with the Business Higher Education Forum, noted that in 2015 there were 2.3 million US job postings asking for data science and analytics skills, an increase from prior years (p. 5).


The demand for work can be expected to grow in areas such as data analysis, software and applications (apps) development, networking and artificial intelligence (AI), as well as designing and production of new intelligent machines, robots and 3D printers. For example, with the greater use of IoT, firms will need to hire more product managers, software developers (including for smart phones), hardware designers, data scientists, user experience designers and sales managers.
Similarly, there is likely to be job growth in “pure” digital firms. For example, in the United States, the number of employees in e-commerce firms that do not have a physical retail shop rose by 66 per cent between 2010 and 2014, from 130,000 to 210,000.

The UNCTAD report goes on to state, “Whatever the rate of change or ultimate outcome of the process of digitalization, the workers of tomorrow will need skills that enable them to create economic value in a world where many jobs are likely to be replaced by automation, software, AI and robots (Levy and Murnane, 2013). Workers will need to be “racing with the machines” rather than “against them”, finding ways in which their skills complement the tasks that machines can carry out and enable them to use and/or augment AI (p. 65).
Finally, as of September, 2018, the position of Data Scientist holds the number 1 spot of best jobs on Glassdoor, as determined by combining three factors: number of job openings, salary, and overall job satisfaction rating.

In summary, the likelihood of an upward continued hiring trend for data science positions is high.

**Is Data Science Exposure a Differentiator for Job Applicants?**

While a numerical analysis is difficult to provide, the implications of various studies suggest that data science concept familiarity is currently a differentiator for job applicants and is trending toward becoming a base requirement in a growing number of career fields.

Some of these analyst positions are full-time data-science roles as noted earlier in this document. But many other analyst roles are more akin to functional analysts.

The report, *Investing in America’s Data Science Talent: The Case for Action*, arranges the spectrum of data science-related job families as follows (p. 6-10):

- **Analytics-enabled jobs**, including data-driven decision makers and functional analysts. Over 60% of the annual data science-related US job postings fall into the category of analytics-enabled jobs.
- **Data science jobs**, including data engineers, data analysts, data scientists and advanced analysts. These are the full-time data science positions, delineated in the early pages of this analysis as Data/BI Analyst, Data Engineer, Research Scientist.

Evaluating the analytics-enabled jobs, it is clear that a growing skillset of data competency is a desirable factor in hiring decisions. It is not a stretch to predict that therefore the channel of entry-level positions will increasingly demand such skills to support and be ready for data-enabled jobs. For example, when personal computers began to integrate into the offices of
accounting and other company departments, the use of personal computers became an entry-
level requirement for nearly all jobs. Today, grade schools in the United States teach computer
skills.

Data-driven decision makers leverage data to inform strategic and operational decisions.
Specific jobs include Chief Executive Officer, Chief Data Officer, Chief Information Officer,
Director of IT, Financial manager, Human Resources manager, and Marketing manager. Skills
that gain an applicant a salary premium include:

- Business intelligence
- Business solutions
- Cloud solutions
- Data warehousing
- Java
- Product development
- Product management
- Risk management
- SAP
- Software development principles

*(Investing, p6)*

Functional analysts use data and models to inform domain-specific decisions. Specific jobs
include Actuary, Business/Management analyst, Compensation/Benefits analyst, Financial
analyst, Geographer/GIS specialist, HRIS analyst, Operations analyst, Researcher/Research
associate, and so on. Skills that gain an applicant a salary premium include:

- Business development
- Business intelligence and data visualization
- Business process and analysis
- Customer relationship management (CRM)
- Data warehousing
- Microsoft development tools
- Risk management
- Software development principles SQL
- System design and implementation

*(Investing, p6)*

This information, taken together with the trend data from the previous question, provides
confidence that data science exposure is very likely a differentiator for job applicants, and in the
future will be a necessity for job applicants. While the job titles may change and focus areas
within the broader data science skillset may shift, the basic skills look to be in demand for the
foreseeable future.
Curricular and Programmatic Initiatives

The growth of data from all sectors shows no signs of slowing, and studies affirm that it seems to be accelerating. Understanding and keeping up with our rapidly changing data landscape is increasingly challenging. For those in higher education, today’s challenge is clear – to develop students’ knowledge, skills and abilities so that they are equipped to participate in today’s data-rich world. For those in Jesuit higher education, our challenge is to go one step further – to develop students’ gateway skills to full participation in the workforce and civic engagement in the 21st century. The proposed enhancement of Seattle University’s curriculum with Applied Data Science offerings demonstrates the University’s commitment to a curriculum for a changing world.

Numerous opportunities to enhance the undergraduate and graduate curricular landscape with Applied Data Science offerings exist. An environmental scan of the University Core curriculum revealed several current courses being offered across the University that might be relevant to joint programs in applied data science or data analytics, and graduate degree programs that have initiated Data Science courses and program offerings. As our environmental scan of the market affirmed, increasing exposure to Data Science promises to be a win-win for undergraduates and graduates alike. Recommendations to enhance the curricular landscape and increase exposure to Data Science are a mix of the practical and the aspirational, including some that are do-able in the short-or-near term, and others that will require longer-term investment in planning.

Core Concepts: Data Literacy, Data Science, and Applied Data Science

Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner. The manipulation of data occurs in daily processes across all sectors and disciplines. Data literacy is widely recognized as an essential ability required in today’s global knowledge-based economy. It follows that the strategic value of data literacy education for undergraduates and graduate students alike, is increasingly clear. An understanding of how decisions are informed by data, and how to collect, manage, evaluate, and apply this data in support of evidence-based decision-making, will benefit Seattle University students while they are attending the University, and will increasingly be required in knowledge economy jobs.

For the purposes of this discussion of curricular and programmatic initiatives, the emerging discipline of data science is distinguished from applied data science. At its simplest, data science is a broad umbrella term whereby the scientific method, math, statistics and a whole host of other tools are applied to data sets in order to extract knowledge and insight from said data. Writing for Forbes in 2013, Gil Press describes data science as the “coupling of the mature discipline of statistics with a very young one—computer science.” He continues “The term “Data Science” has emerged only recently to specifically designate a new profession that is expected to make sense of the vast stores of big data. But making sense of data has a long
history and has been discussed by scientists, statisticians, librarians, computer scientists and others for years.\(^1\)

Essentially, Applied Data Science is a specialization (with a small “s”) within the emerging discipline of Data Science. Applied Data Science appeals to undergraduates and graduates who want to acquire practical skills for real world data problems, as well as to students interested in pursuing a career in Data Science. Students taking Applied Data Science courses will develop foundational Data Science skills to prepare them for a career or further learning that involves more advanced topics in Data Science, including big data, artificial intelligence (AI), and deep learning. The specialization entails understanding what is Data Science and the various kinds of activities that a Data Scientist performs. It generally familiarizes learners with various open source tools used by Data Scientists and teaches students about the methodology involved in tackling data science problems. The specialization also provides knowledge of relational database concepts and the use of various standardized languages to access and manipulate databases. Providing opportunities for students to apply their newly acquired skills and knowledge, lectures, labs, and projects provide hands-on experience tackling interesting data problems.

**Campus Environmental Scan**

This campus environmental scan focuses on the University Core Curriculum and data literacy education, joint academic programs in Applied Data Science, and the graduate curriculum. The scan begins with a recommendation to establish a “data intensive” course designation similar to the “writing intensive” designation used at Seattle University.

Currently, there is no indication that a course at any level is data intensive. Undergraduate and graduate students seeking to choose the most data intensive coursework possible are on their own – left to read and assess course descriptions, confer with their advisors, check in with their peers, or otherwise try to make sense of the academic catalog.

Just as we have a system for designating writing intensive courses, so, too, can we establish a similar mechanism for identifying data intensive courses. The Applied Data Science Center discussed later in this report could be given responsibility for managing such a process. Specific policies for designating courses as data intensive courses should be developed by the Center, and should include requirements of substantial hands-on work with real-world data, whether in the context of homework, lab assignments, projects, or other avenues.

**University Core Curriculum and Data Literacy Education**

Given the role of the University Core, and its relationship to students’ major classes and out-of-class experiences, focusing on the Core Curriculum provides a natural starting point for an exploration of data literacy education for undergraduates at Seattle University. The “Welcome to the University Core” describes its centrality to academic affairs –

\(^1\) https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#72c3b4855cfc
The Core Curriculum is Seattle University’s common undergraduate educational experience. Indeed, it is the very heart of your Seattle University education, embodying the university’s mission and helping you develop as leaders for a just and humane world. The Core isn’t simply a distribution requirement intended to make you more “well rounded.” Instead, it is a thoughtfully designed, integrated curriculum created to help all SU students grow as scholars, as citizens, and as reflective and engaged whole persons, ready to make a difference in the world. ([https://www.seattleu.edu/core/](https://www.seattleu.edu/core/))

### Joint Academic Programs in Applied Data Science

Current courses being offered across the University that might be relevant to joint programs in applied data science or data analytics were researched. Most of the courses are found in the Albers School of Business & Economics (business analytics) as well as the College of Science & Engineering (math and computer science), but a sprinkling of courses being offered in Arts & Sciences was also identified.

The courses identified are listed below, separated by college and program. The list that follows represents a snapshot at a specific moment in time - ~Spring 2018; it is not intended to be exhaustive. Additions are welcome.

- **Albers School of Business and Economics**
  - **Business Analytics**
    - BUAN 3210 Data Visualization and Communication
    - BUAN 3280 Law and Ethics for Business Analytics
    - BUAN 4210 Programming and Data Management for Business Analytics
    - BUAN 4310 Data Mining and Big Data Analytics
    - BUAN 5210 Communicating and Visualizing Data
    - BUAN 5260 Mathematical Models for Decision-Making
    - BUAN 5281 Law for Business Analytics
    - BUAN 5310 Statistical Learning for Business

  - **Economics**
    - ECON 2100 Business Statistics
    - ECON 3100 Quantitative Methods and Applications
    - ECON 4110 Applied Econometrics
    - ECON 4120 Forecasting Business Conditions
    - ECON 5000 Introductory Business Statistics
    - ECON 5100 Statistical Applications and Quantitative Methods
<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECON 5300</td>
<td>Applied Econometrics</td>
</tr>
<tr>
<td>ECON 5305</td>
<td>Economics and Business Forecasting</td>
</tr>
</tbody>
</table>

- **Executive Leadership MBA**
  - EMBA 5310 Statistical Modeling for Executives

- **Information Systems**
  - IS 3150 Introduction to Information Systems
  - IS 3200 Fundamentals of Programming in Business
  - IS 4660 Database Systems in Business
  - IS 5201 Programming for Business Analytics
  - IS 5305 Data Management In Business
  - IS 5310 Data Mining for Business Intelligence
  - IS 5315 Big Data Analytics

- **Sport Business Leadership**
  - SADL 5170 Measurement and Evaluation in Sport Management

- **College of Arts & Sciences**
  - **Anthropology, Sociology, and Social Work**
    - SOCW 4010 Social Work Data Analysis
    - SOCW 5410 Social Work Research I: Methods and Design
    - SOCL 3130 Sociological Research Methods
  - **Digital Design**
    - ARTD 4760 Digital Media III: Interactive Graphics
  - **Communication**
    - CMME 2200 Multimedia Reporting
  - **Criminal Justice**
    - CRJS 3020 Criminal Justice Statistics
    - CRJS 5070/5080 Statistical Analysis w/ Lab
    - CRJS 5240 Crime Mapping
- **Environmental Studies**
  - EVST 3400  Research Design and Statistics
  - EVST 3500  Intro to Geographic Information Systems

- **Institute of Public Service**
  - PUBA 4400  Research Design and Statistics
  - PUBM 5440  Applied Data Analysis
  - PUBM 5450  GIS for Public Administrators

- **Psychology**
  - PSYC 3030  Statistics and Research Methods I
  - PSYC 3050  Statistics and Research Methods II
  - PSYC 4030  Advanced Statistics and Experimental Design

- **School of Law**
  - INTP 393  Big Data, Informed Consent, And The Future Of Privacy Law
  - INTP 320  Copyright Law

- **College of Science & Engineering**
  - **Chemistry**
    - CHEM 3000  Quantitative Analysis
  - **Computer Science**
    - CPSC 1220  Data-Driven Prob. Solving & Programming
    - CPSC 2430  Data Structures
    - CPSC 3300  Fundamental of Databases
    - CPSC 4100  Design and Analysis of Algorithms
    - CPSC 4910  Data Science
    - CPSC 4910  Cloud Computing
    - CPSC 4910  Artificial Intelligence
  - **Electrical & Computer Engineering**
    - ECEGR 1000  Computing for Engineering
    - ECEGR 1200  Digital Operations and Computation
Graduate Curriculum

Seattle University currently offers 176 graduate certificate and degree programs, which are generally decentralized and offered through seven of the nine Colleges and Schools within the University. While several programs—most notably in Albers School of Business & Economics and the College of Science & Engineering, have already initiated robust Data Science courses and program offerings, most graduate programs—most notably in the Colleges of Arts & Sciences, Education, and Nursing, have limited exposure to the discipline. This is despite a general sense among members of these programs that increased awareness and utilization of Applied Data Science would generally enhance students’ educational experience. In these programs, Data Science is generally regarded as “highly beneficial” but not “essential”.

Seattle University graduate programs that have already identified elements of Applied Data Science as “essential” to their program learning outcomes (e.g., Master of Science in Business Analytics; Master of Science in Computer Science) have already developed program-specific data science coursework and faculty expertise. These programs and course offerings can serve
as useful templates for other programs interested in increasing their own Data Science capacity. University programs have been established to provide students who have little-to-no previous programming experience the foundational tools necessary to develop a relative level of Data Science proficiency (e.g., Graduate Certificate in Computer Science Fundamentals); however, graduate students in programs outside Science & Engineering have limited exposure to these course offerings.

The inclusion of discipline-specific coursework that addresses real-world problems involving large dataset collection, cleaning, manipulation, analysis, and reporting is of broad interest to a substantial portion of our graduate students and programs. There are notable opportunities in several of our programs that do not currently incorporate Applied Data Science to benefit from increased curricular offerings and interdepartmental, campus wide expertise. As we begin to develop and propose additional course content, it will be important to ensure courses and resources are able to “fit” within each interested programs’ curricular focus and credit requirements expectations. Expanding inter-departmental opportunities for, and resources in, Applied Data Science teaching and learning could benefit multiple disciplines, particularly those who currently have limited offerings or expertise.

Recommendations

**University Core Curriculum and Data Literacy Education**

Leveraging intersections at the upper and lower ends of the University Core promise the most immediate and perhaps the most significant benefits for faculty and undergraduates alike. Supporting recommendations (#1 and 2) are accompanied by two additional recommendations (#3 and 4) that consider curricular and/or other academic options outside the Core. Two scenarios – building data literacy education into existing majors and pursuing an additional requirement as a means of advancing data literacy, were not fully explored by the Subcommittee.

**Recommendation #1:** Leverage the upper end of the University Core by focusing on the 3000 level, developing a suite of “3200” level courses, and/or by focusing on the Global Challenges courses.

**Recommendation #2:** Expand the lower end of the University Core by taking advantage of existing processes (mechanics) for Core expansion and development.

**Recommendation #3:** In recognition of the fact that data literacy education can also be developed, introduced and sustained by the provision of courses and/or tracks outside of the University Core, exploration of the provision of courses and/or tracks through the “Center” or through New and Continuing Studies is recommended.

**Recommendation #4:** Exploration of the potential of data literacy or data science boot camps or workshops is also recommended. Providing immersive, hands-on, short-duration experience, boot camps could be sponsored by the “Center,” New and Continuing Studies or
other schools and/or colleges within Academic Affairs.

**Joint Academic Programs in Applied Data Science**

**Recommendation #1:** The Subcommittee recommends that Seattle University develop an interdisciplinary undergraduate minor in applied data science. This minor should include foundations in computer programming, databases, statistics, and data visualization.

**Recommendation #2:** The Subcommittee recommends exploring the possibility of a graduate certificate or master’s program in data science. This should be done in consultation with whatever college(s) / department(s) would host such a program (such as Mathematics or Computer Science in the College of Science and Engineering).

**Recommendation #3:** The Subcommittee recommends that any development of graduate programs in data science consider the feasibility and potential demand for dual degree programs with data science and existing graduate programs such as criminal justice or business analytics.

**Graduate Curriculum**

The following three recommendations are based on understanding of the current Seattle University graduate program offerings and the value in considering Applied Data Science accessibility across disciplines

**Recommendation #1:** Developing of an interdisciplinary, comprehensive, stand-alone *Introduction to Data Science* course for Seattle University graduate students who have no programming background.

**Recommendation #2:** Develop a “Programming Boot Camp” aimed at students who have no prior programming experience, but who will benefit in their studies and work from developing a knowledge of a programming language such as Python. The Computer Science department currently offers a programming boot camp ([https://www.seattleu.edu/scieng/computer-science/certificates/cs-fundamentals/programming-boot-camp/](https://www.seattleu.edu/scieng/computer-science/certificates/cs-fundamentals/programming-boot-camp/) accessed 25 September 2018)) aimed at students intending to do intensive software development. There is need for a similar program for students who will not be going into computer science, but who do need a working familiarity with a high-level programming language. The structure and details of the new boot camp should be determined by a set of representatives of graduate programs that would subscribe to the boot camp.

**Recommendation #3:** Ensuring students in all Seattle University graduate programs have the necessary campus resources for assisting them with large dataset collection, analysis, and dissemination.
Although each program, if not each graduate student, has unique educational needs and professional goals, these three recommendations are intended to positively impact the broad swath of students while not instituting new program requirements. The first two recommendations are intended to be revenue neutral, if not revenue-generating, and the third recommendation is intended to ensure a sufficient Data Science resource assessment is being accomplished as part of the standard program review process.

Roadmaps

The major steps for achieving the various recommendations follow. Also included are select milestones needed to reach these goals.

University Core Curriculum and Data Literacy Education

**Recommendation #1:** Leverage the upper end of the University Core by focusing on the 3000 level, developing a suite of “3200” level courses, and/or by focusing on the Global Challenges courses.

Focusing on the upper end of the University Core, the 3000 level has the most promise. This has the potential to draw transfer students as well as upper level juniors. The 3000 level also promises to draw students who are more motivated, able to apply concepts, skills and abilities to their majors. Some form of computing background is desirable and will ensure academic success. There are various means to achieve this prior to starting the course itself, including pre-requisites in specific disciplines, online self-paced workshops or courses, boot camps etc. Pre-screening may also be used to determine students’ coding skills etc. “Entering expectations” (knowledge, skills, abilities etc.) would also be reflected on the online course description, syllabi etc.

Development of a suite of “3200” level courses provide another gateway into the upper end of the University Core. All the classes here would be computing and serve as a bridge to help students move into other classes.

A third avenue into the upper end of the University Core is provided by the Global Challenges courses. (Humanities, Social Sciences, Natural Sciences) (UCOR 3400-3440; 3610-3640, 3810-3840). Every student must select two (2) courses of the three (3) that are most unlike their major. This win-win: provides faculty with the opportunity to develop courses in their disciplines. Several enabling strategies are suggested – 1) make the course(s) as flexible as possible up front, i.e. design the course(s) to be data and topic agnostic so as to work for students taking courses “outside” of their major area; 2) frame the course(s) to align with the global challenge(s), presenting the challenge(s) broadly at first, i.e. start small, carefully constructing the learning objectives and the syllabus/I; 3) Leverage the value and benefit of cross-listing the course(s).

Regardless of the path(s) chosen to integrate data literacy education into the existing University Core at the upper end, promotion using a variety of methods and modes is
key to the successful launch of data literacy education, and its continuing vitality overall.

While more classes in the lower Core is good for data literacy, we do not recommend counting on any lower level Core courses as prerequisites to other courses.

**Recommendation #2:** Expand the lower end of the University Core by taking advantage of existing processes (mechanics) for Core expansion and development.

We recommend working with the Director of the University Core Curriculum to identify the major steps needed to expand the lower end of the University Core.

**Joint Academic Programs in Applied Data Science**

**Recommendation #1:** The committee recommends that Seattle University develop an interdisciplinary undergraduate minor in applied data science. This minor should include foundations in computer programming, databases, statistics, and data visualization.

The Applied Data Science Minor is a six-course (30 credit) sequence designed to provide a foundational core of skills related to data science as well as subject-based electives to help students apply their learning to their individual disciplines.

The Applied Data Science Minor is designed to insure the following Program Learning Outcomes (PLOs).

1. Students model and manage data using professional tools and concepts.
2. Students write code that accesses, manipulates, and quantifies data in order to fulfill their requirements.
3. Students apply principles of statistics and mathematical analysis to data in order to gain understanding or otherwise interpret information.
4. Students produce well-reasoned and informed analysis of data suitable for their given audience and discipline.
5. Students utilize data science in a way that is commensurate with professional practice within their chosen subject area.

In order to complete the Applied Data Science Minor, students must complete six courses. These courses are divided into four “Foundation Courses” and two “Subject-based Electives.” The Foundation Courses are meant to provide a solid understanding of core concepts and techniques used in the field of Applied Data Science. These courses especially address PLOs One through Four. The Subject-based Electives could address any of the PLOs, but they are expected to focus on PLO Five.

- Something to fill a databases requirement (PLO 1)
o CPSC 3300 - Fundamentals of Databases

- Something to fill a programming requirement (PLO 2)
  o CPSC 1220 - Data Driven Problem Solving and Programming (Python)

- Something to fill a stats/modeling requirement (PLO 3)
  o Future MATH course

- Something to fill a visualization / analysis requirement (PLO 4)
  o Exploring with Albers, Arts and Sciences, New and Continuing Studies

Individual departments and programs are encouraged to think about how data science impacts their field and propose courses to fulfill this aspect of the Applied Data Science Minor. Pre-approved courses can be chosen by students, and other courses should be allowable via petition. (It will be much easier if programs get their courses pre-approved.)

The goal of these courses is to allow each program to shape the Data Science Minor to most benefit their students.

**Resource Requirements**

- Administrative support for the college/department that houses this minor
- Work units for MATH to develop and deliver new course
- Resource requirements for MATH, CPSC, and other departments serving the foundation courses

**Recommendation #2:** The committee recommends exploring the possibility of a graduate certificate or master’s program in data science. This should be done in consultation with whatever college(s) / department(s) would host such a program (such as Mathematics or Computer Science in the College of Science and Engineering).

**Recommendation #3:** The committee recommends that any development of graduate programs in data science consider the feasibility and potential demand for dual degree programs with data science and existing graduate programs such as criminal justice or business analytics.

These ideas have not yet been fleshed-out, as they will require substantial conversation with the college(s) and department(s) that would potentially house such programs. There is potential for such programs to exist synergistically with existing graduate programs, offering dual degree programs in, for example, data science and business analytics, or data science and criminal justice.
Market research for graduate programs should also investigate whether demand is highest for an online program, a face-to-face program, or a hybrid, and what the pedagogical implications and content demands for each would be. The newly-developed online MSBA program in Albers has found that there are many ways in which their program will differ from the face-to-face MSBA program.

**Graduate Curriculum**

**Recommendation #1:** Developing of an interdisciplinary, comprehensive, stand-alone Introduction to Data Science elective graduate course for Seattle University students who have no programming background.

There appears to be substantial interest and market potential in offering graduate students a rigorous Data Science survey course intended to (1) introduce students to the discipline and (2) provide a baseline level of coding competency to which students can apply their knowledge in pursuit of their scholarly work and postgraduate career prospects. An outline of potential course topics to be presented have been created (Table 1).

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data Science and its real-world applications&lt;br&gt;Introduction to programming language to be used</td>
</tr>
<tr>
<td>2</td>
<td>Using and manipulating variables&lt;br&gt;Reading and saving datasets into a file</td>
</tr>
<tr>
<td>3</td>
<td>Understanding common data structures and data types</td>
</tr>
<tr>
<td>4</td>
<td>Analyzing and interpreting information in complex datasets</td>
</tr>
<tr>
<td>5</td>
<td>Data cleaning and error handling</td>
</tr>
<tr>
<td>6</td>
<td>Modifying, searching, and sorting lists and arrays</td>
</tr>
<tr>
<td>7</td>
<td>Using Boolean logic, if statements, loops, and control structures</td>
</tr>
<tr>
<td>8</td>
<td>Creating functions and methods</td>
</tr>
<tr>
<td>9</td>
<td>Conducting algorithmic problem solving and hypothesis testing</td>
</tr>
<tr>
<td>10</td>
<td>Data visualization and reporting</td>
</tr>
</tbody>
</table>

*Tentative topics list and sequence. To be further developed and revised in consultation with potential course instructors and graduate program faculty representatives.

This topics list would be further developed and refined by presumptive course instructors and in consultation with faculty representing the programs likely to have student enrollees. The target Seattle University graduate student audience for this course would
be those who have an interest in learning the applications of coding relevant to their program of study but have little-to-no coding or data science experience.

While much of the learning content covered in this proposed course is available elsewhere (e.g., online via Codecademy, DataCamp, et al.), many of our students, particularly in the humanities and social sciences, appear hesitant to utilize them. Those online self-paced courses tend to lack the personal student-instructor interaction that attracts many of our graduate students to Seattle University. Providing an on-campus course, taught by our faculty and with an understanding of the applications of data science of most interest to our students could have much potential.

Offering a single course that is available to all graduate students\(^2\) appears to be a more cost-effective approach than promoting multiple discipline-specific courses. The course instructor would be encouraged to incorporate discipline-specific examples that are most relevant to the student enrollees. Examples of intended course features and possible optional features are presented in Tables 2 and 3, respectively. If a program demonstrates sufficient demand to run or develop their own course, the topics list (Table 1) could serve as a generic template for them as well.

<table>
<thead>
<tr>
<th>Table 2. Intended course features for a proposed Introduction to Applied Data Science course.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Course Feature</strong></td>
</tr>
<tr>
<td>Credit-bearing course</td>
</tr>
<tr>
<td>Instructed by Seattle University faculty</td>
</tr>
<tr>
<td>3 credits</td>
</tr>
<tr>
<td>Offered as an interdisciplinary elective</td>
</tr>
<tr>
<td>No prior programming experience</td>
</tr>
<tr>
<td>Covers most common topics of a Data Science pertinent to non-computer science subject matter experts</td>
</tr>
</tbody>
</table>

\(^2\) The proposed course is intended for programs not currently incorporating Applied Data Science into their curricula. It does not replace currently offered program-specific courses and is not designed to meet any current program major requirements.
Includes hands-on, practical application of Data Science to real-world scenarios

Students should be able to develop a level of comfort applying what they learn to their graduate program curricula, scholarly projects, and postgraduate career prospects.

Scheduled at a convenient time

Scheduling at a time not used by most programs increases potential student enrollee pool. A suggested time is once per week during a currently unused scheduling block (e.g., Friday late afternoons).

Develop proficiency in a programming language most commonly used in Data Science

Teaching in one language for the entire course increases students' depth of coding understanding and language fluency. Although the specific language will be selected by the instructor, it will be particularly relevant to Data Science end-users in various vocations.

### Table 3 Additional course options for consideration in a proposed Introduction to Applied Data Science course.

<table>
<thead>
<tr>
<th>Course Feature</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initially offer as a Special Topics Elective course</td>
<td>Allows course to be piloted, revised, and honed prior to going thought the full course approval process. Permits course to be advertised and taught during the coming academic year (e.g., Spring 2019).</td>
</tr>
<tr>
<td>Cross-list as 4000-level undergraduate course</td>
<td>Increases potential student enrollee pool. Provides undergraduates an opportunity to experience graduate education.</td>
</tr>
<tr>
<td>Provided as online course with in-person study groups</td>
<td>Increases student enrollee pool. Providing in-class study groups may help alleviate potential complications of online-only course delivery.</td>
</tr>
<tr>
<td>Incorporate program-specific tie-ins</td>
<td>If the course is heavily subscribed by students in a particular graduate program, specific tie-ins to their use of Data Science may increase student engagement and appreciation of the content's utilization. If the course is heavily subscribed, it may be possible to offer sections tailored to the students’ programs of study.</td>
</tr>
</tbody>
</table>

The next steps advancing Recommendation #1:

- Identify likely faculty and obtain provisional approval from their dean to submit a Special Topics Elective course proposal, with expected course stipulations (e.g., minimum enrollment, workload allocation, etc.) (Summer/Fall 2019)

- Survey interested graduate program faculty and students to ensure maximum feasibility and suitability of the course (e.g., selection of topics, scheduling, course format, etc.) (Fall 2019/Winter 2020)

- Submit a Special Topics Elective course proposal. Advertise among targeted graduate programs and student enrollees. (Fall 2019/Winter 2020)

- Conduct course. Review student feedback (Spring 2020)
• Determine potential course revisions. Consider subsequent offerings, scheduling, and conversion from a Special Topics course to a permanently numbered course. (Summer/Fall 2020)

**Recommendation #2:** Developing a “Programming Boot Camp” similar to the Computer Science Department’s existing boot camp, but targeted to the needs of non-computer science graduate students.

The College of Science & Engineering currently offers a 3-course (9 credit) summer intensive “Programming Bootcamp” (i.e., CSPC 5001, CSPC 5002, & CSPC 5003). The course sequence is an accelerated introduction to computer programming designed for students who have limited-to-no prior programming experience and is part of the Graduate Certificate in Computer Science Fundamentals (Career Change Certificate) program. While this sequence is not at an appropriate level for non-computer science students, it can serve as a model for a program aimed at non-computer science students, emphasizing the programming skills that these students will need, in a high-level language such as Python.

Developing a Programming Boot Camp will ensure students in other Seattle University graduate programs who are interested in a comprehensive introduction to Data Science have access to this level of skill development. Multiple features of the Boot Camp model make it particularly promising and generally compatible with other Seattle University graduate programs, including: (1) The existing programming boot camp has been well received; (2) it requires minimal prerequisite knowledge; and, (3) courses are taught during Summer term.

The next steps advancing Recommendation #2:

• Meet with the Computer Science department to discuss a Programming Boot Camp with appropriate content for non-Computer Science graduate students. (Spring 2019)

• Offer Boot Camp faculty the opportunity to meet with interested student enrollees and graduate faculty from other programs to solicit input and gauge level of interest. (Summer/Fall 2019)

• Determine appropriate curricula for a non-computer science student Programming Boot Camp. (Fall 2019/Winter 2020)

• Promote the Boot Camp broadly to graduate students and graduate program faculty across campus. (Winter/Spring 2020)
Recommendation #3: Ensuring students in all Seattle University graduate programs have the necessary campus resources for assisting them with large dataset collection, analysis, and dissemination.

Applied Data Science resources include providing students’ and faculty access to the hardware, software, storage, and knowledgebase necessary to collect, store, analyze, and disseminate course assignments, scholarly works, and capstone projects. As described in Part 1 of this report, there is currently a divergence of Data Science curricular needs and utilization among our graduate programs, which appear to be currently addressed within each College or at the individual program level. The prospect of additional University-wide Data Science resources is the focus of another Taskforce Subcommittee and we defer to their recommendations.

A major challenge we encountered in information gathering for this report was identifying the most pertinent faculty representatives from each program. Having a current database of faculty and administrator contacts who could speak to Data Science programmatic needs, opportunities, and interests would be highly beneficial.

The next step advancing this recommendation is to:

- Work with associated deans and program administrators to establish an Applied Data Science faculty contact list representing all interested graduate programs. (Summer/Fall 2019)
- If and when a graduate program in data science is developed, consult with existing graduate programs to explore the feasibility of offering dual degree programs with data science. The same could be done with existing applied data science graduate programs such as the MS in Business Analytics and MS in Computer Science.

Challenges

A summary of the core challenges of successfully implementing data science across the undergraduate and graduate spectrum at Seattle University follows.

A solid foundation in data literacy has become vital across disciplines. Seattle University must ensure that our students develop the foundational data literacy skills that they will use throughout their academic, personal and professional lives. And in a growing number of fields, there is a need for not just data literacy, but applied data science skills as well. There are many strategies and best practices for both implementing applied data science and for teaching data literacy. A strategic review of best practices will inform assessment of the various data literacy and data science education options. Particularly relevant to our needs are best practices for the appropriate timing of this education; delivery (courses vs. workshops etc.); emerging teaching approaches and learning environments; engaging content with real world data to foster
innovation; successive, iterative, practical hands-on learning; and best practices for assessment and evaluation of data literacy and applied data science skills.

Focusing on the undergraduate curriculum, there needs to be agreement on what elements of data science are necessary in undergraduate curricula. Identification and synthesis of the skills and abilities that together comprise various levels of data literacy and data science is an important first step for assessment of the costs/benefits associated with the various avenues for data science at Seattle University.

While there are notable challenges in attempting to propose graduate-level Applied Data Science teaching and research as part of a University-wide initiative, the field has the potential to be a unifying and cross-discipline curricular element, akin to math, writing, or research methods courses that are included in most of our programs. Like courses in statistics or writing, coursework in Applied Data Science should look different when targeted to students in specific disciplines and taught by those faculty through the lens of their vocational or research specialization. However, many of the graduate programs within Seattle University that do not currently offer Applied Data Science coursework but could benefit from it lack the faculty expertise or bandwidth to do so. Incentivizing the faculty who will be involved in developing pedagogy in these areas will require both fiscal incentives as well as an investment in training and professional development.

Additionally, a common concern voiced by students has been the impact that additional coursework, even if beneficial to their program of study, would increase credit expectations or their time to degree completion. An additional concern, which was illustrated in a survey of a social science graduate program cohort (Table 6), was the prospect of taking Applied Data Science coursework that includes computer programming as prerequisite knowledge. This survey showed that students overwhelmingly perceive a course in Applied Data Science would be useful for their current program of study.

| Table 6. Survey result of a first-year social science graduate student cohort asked about the utility and applicability of Applied Data Science* to their anticipated program of study. |
|-------------------------------------------------|---|---|---|
| Would a course in Applied Data Science be useful for your intended program of study and/or scholarly project? | Yes | No | Uncertain |
| | 72% | 5% | 23% |
| | (n = 28) | (n = 2) | (n = 9) |
| Are you relatively proficient in one or more commonly used Data Science programming languages (e.g., Python, R, SQL, Matlab, C++) | 3% | 97% | - |
| | (n = 1) | (n = 38) | |
| Would you only be interested in taking an Applied Data Science course if it did not require prior computer programming experience? | 44% | 3% | 54% |
| | (n = 17) | (n = 1) | (n = 21) |

*The terms “Applied Data Science” and “relatively proficient” were defined to ensure common understanding among respondents.

Surveying graduate program representatives illuminated a divergence of current interest among faculty and students surrounding the potential of including Applied Data Science course
offerings in graduate programs where it is currently not part of the curriculum. A more in-depth analysis (i.e., program-by-program review, Recommendation #3) would be beneficial to determine the potential role Applied Data Science coursework and learning objectives could play in achieving programs’ overall learning objectives and intended outcomes.

Major challenges to considering campus-wide recommendations for providing and enhancing Applied Data Science opportunities at Seattle University include the professional specialization of each degree and certificate as well as the diversity of intended program outcomes among our active graduate programs. The lack of University-wide graduate program course requirements (cf. the undergraduate University Core curriculum) and the generally autonomous nature of the University’s graduate programs further complicate University-wide offerings. Nonetheless, there is a perceived market demand and enhanced postgraduate employability for students across disciplines with Applied Data Science skills.
The Case for an Applied Data Science Center

This subcommittee’s primary and most significant recommendation is the establishment and maintenance of an Applied Data Science Center (“Center”) for Seattle University. We found virtually all aspects of Applied Data Science work for students, faculty, staff, and alumni at the University would be most efficiently executed through the resources and functions that a Center would provide. In addition, the Center should serve as a focal point for connections between Seattle University and the broader community in the realm of data science and the application of this technology to our lives.

The four key operation principles of the Center should be: (1) Resource Hub; (2) Best Practices Center; (3) Dedicated Staffing; and (4) Event Coordination. While it is difficult to be precise about the exact structure of the University Center given the pace of change in applied data science, we believe that continued refinement of our operating principles with more details about technological infrastructure will allow the Center to take shape based on its function.

Discussion

1. The Resource Hub

The Center should be a Resource Hub focusing on Applied Data Science learning and innovation for Seattle University. Leading as a peer-among-peers, the Center would:

- Help guide professional development of students, faculty, and staff
- Help establish new data science curriculum across campus
- Find and license necessary and appropriate large data sets
- Provide data science consulting for faculty and student research
- Facilitate interdisciplinary collaboration on data science related scholarship
- Coordinate with ITS and the Library to ensure access to relevant technology
- Guide faculty/students through the labyrinth of choices regarding curriculum and research
- Provide intellectual property clearance and counseling
- Promote continuing emerging technology
- Serve as a hub for student and academic projects.
- Be a thought leader on the ethical use of data analytics (including identifying the privacy risks specific uses of analytics pose to privacy and pioneering responsible processes that should accompany any such uses)
- Provide a resource for reflective engagement about what data science means for the individual and our culture

It is important that a focal point exists for this information. Consequently, the Resource Hub needs to be an information clearing house for University-wide technology and data literacy in all disciplines.

Additionally, the Resource Hub can ensure that some specific essential infrastructure decisions are made thoughtfully recognizing their resource requirements, including:
• An emphasis on general technical support for Faculty and Students through integration with existing ITS infrastructure
• Usage data and infrastructure IO (e.g. asset tracking, budget, contracts, etc.)
• Purchasing and license administration for campus wide central data science solutions

Example: A faculty member could seek help from the Center with how to set-up an AWS, Microsoft Azure, Google Cloud, etc. Account at the best available pricing.

Data science infrastructure is essentially a set of tools (AWS, Microsoft Azure, Matlab, Atlas.ti, etc.), and support for these and other related well-known technologies. The Center will identify such tools that many faculty and/or students at Seattle University either currently use or would like to have available, and coordinate University-wide licensing for these tools, providing a more cost-effective and administratively simpler solution for the University than managing myriad individual licenses. However, not all problems need just one tool. The support of experimental problems requires collaboration rather than simply tool dissemination. This in turn requires robust governance about what the Center can support (e.g. PyTorch – a python package for tensor computation and deep neural networks; Docker – application containerization, etc.). Driving the decision-making process on what to support should be a major task for the office of a Director of Applied Data Science in close collaboration with IT. Setting up governance at the Center that has equal voice for conservative central data science solutions, as well as giving opportunity for creative experimental solutions, should provide the tension necessary for balanced and appropriate solutions for the University.

We found many examples of specific infrastructure needs at Seattle University. Briefly, we have identified two representative examples of centralized infrastructure, (AWS and MatLab) and a creative experimental example (genomics and post-genomics data processing). Hopefully, a picture of how the Center could address these unique problems helps outline the function, and thus form, of the Center.

Centralized Infrastructure Example: There is growing need for institutional access to MatLab. Many professors and students pay out of pocket with personal credit cards for use (especially since current University licensing does not include research). Included in a total academic headcount (TAH) license is the ability to install on local hardware, cloud access, and MatLab’s suite of toolboxes for all faculty/staff/students which would greatly simplify infrastructure demand forecasting. This is a much cheaper option than providing staff support for all other platforms and would help focus the expertise of those that do support faculty to the more creative issues of open-source support. Additionally, it is much easier for students to step down to free open-source languages post-graduation than it is to pay for and learn MatLab while already out of school. A simple MatLab-to-python short course for graduating students could cover the transition easily. Support and license distribution are handled by Mathworks under a TAH subscription. TAH is an annual subscription. The current pricing for the full suite is $22,869 (88 Products included, not including tax).

In addition to the advantages that the Center could provide by obtaining the optimal institutional MatLab licensing, the Center could keep the University updated with real-world solutions that use MatLab.

Our interviews revealed at least twelve faculty members currently needing a MatLab TAH.
*Creative Experimentation Example:* As mentioned, we have also identified an example of the need to support more creative and enabling technology that focuses on open-collaborative code--genomics and the post-genomic sciences.

Natural science education at Seattle University focuses substantial effort toward placing students in advanced degree fields, especially those related to medicine and allied health. Computational skill is important in the training of these students, but not the only important skill. Also of primary importance is the immense amount of information these students must learn about physical reality. Any way faculty can integrate the use of computational tools to help students in their study of medicine provides training in two ways: scientific-inquiry based learning; and hands-on engineering appreciation of technological change.

The sequencing of the human genome has spawned many new data driven approaches to medicine, each with their own unique infrastructure needs. Genomics itself can be supported more straightforwardly than others, as it has been around the longest. Many web-based applications exist that professors are currently using to teach students how to collect, store, and process DNA sequence data. However, these servers are hosted by institutions that limit the number of concurrent logins, so whole classroom crashes are common (e.g. BLAST, CLUSTAL, PRIMER-BLAST, Phylogenetic Tree and Maximum Likelihood Analysis). In addition, using other institutions’ resources limits the ability for Seattle University users to control the tools used and the problem sets. Building these databases on University controlled infrastructure is straightforward and an example of a more individualized infrastructure support that the Center could provide.

The fields of proteomics and metabolomics (and other -omics sciences) continue to be funded to make data processing more user friendly. These sciences deal with massive amounts of data and long computationally intensive simulations. However, the code base for many of these projects are research-focused and are becoming more centrally supported and accessible to all (mainly by large R01 research institutions, e.g. Skyline and MetaboLights). If we are to prepare our students for advanced degrees in these laboratories and clinics, they need to at least understand the ad hoc nature of these sciences, if not be able to have a fundamental working knowledge of them. For the Center, this requires a more flexible approach to support, and a willingness to go above and beyond central solutions to creative collaboration. While we have highlighted our understanding of biochemical medicine here, other fields (such as Astronomy) also deal with a similar daunting landscape of chaotic innovative collaboration. It must also be the duty of the Center to help students and faculty navigate these chaotic waters.

There are also many situations in which faculty or students at Seattle University run into data science related questions in their research. While sometimes it is beneficial for them to have resources to learn the appropriate data science tools to engage with these questions, at other times it is much more efficient for them to consult with a data science expert to address their questions. Currently, no official consulting service exists to help the on-campus community at Seattle University. Much consulting work happens unofficially, where faculty ask around amongst those they know to see if anyone can point them to someone who can help them. The Center will provide a centralized resource for data science consulting, so that all faculty will have a known entity to contact with their data science research questions. At many universities with data science programs, it is common to offer a consulting service in which students provide consulting services alongside faculty or staff experts, allowing the consulting services to also become an educational opportunity for our students.
In addition to faculty/student communications and the Center, there will also be situations in which the role of the Center will be to help put faculty or students into communication with others at Seattle University (or in the broader community we will be in communication with) when there is a known person who could help address their questions. For example, an undergraduate math major was recently involved in summer research regarding modelling of wildfire outbreaks. They wanted to learn more about GIS systems, as this would be beneficial to their research work. They were unaware of the faculty expertise in GIS that currently exists in both the Environmental Science and Criminal Justice programs (although fortunately a faculty member was able to inform the student of these resources). In the future, a student in a similar situation could contact the Center to learn who on campus they could contact who would have relevant expertise to their work.

Lastly, at Seattle University, where our mission is the education of the whole person, the Center’s resource hub should also be a source of relevant knowledge and reflective engagement about the application of technology to our lives. The Center’s mission should include investigation into what these analytics tools mean for our culture, communities, students and the ethics of computation.

For example, one concern is the physical health and well-being of students and the link between computers and a sedentary lifestyle. Data science tools are mainly focused toward an efficiency-driven workplace with most time spent staring at a screen. There is a biochemical and mechanistic link between the dysregulation of the hypothalamic pituitary adrenal axis seen in this technological environment and poor psychiatric and physical health outcomes. This is an emerging global health burden and the resource hub must provide coaching to students and faculty on how to increase computational skill while at the same time becoming happier and healthier individuals.

As another example, there has recently been increased awareness of social justice implications of data science. During fall quarter of 2018 the Seattle University Math Department will be hosting a guest speaker who works with issues of data science & social justice. They will discuss ways in which the field of data science has created uncomfortable or even dangerous situations for transgender individuals, using algorithms that do not adequately consider many of the nuanced issues of gender identity. As we come to rely on computational tools to inform us about decisions with outcomes that should be carefully weighed by humans, we must work hard to understand the hidden relationship between the algorithm and social justice. In the future, event such as these could be coordinated through the Center to help lead a campus-wide conversation on these issues.

2. Best Practices Center

The Center should be the University’s central repository of best practices on interdisciplinary approaches to working on applications of data science, including use of data in all aspects of academic life at Seattle University. By being the curator of best practices, the Center would be in a privileged position to collect experiences and provide thought leadership on the ethical, legal and practical issues implicated in Applied Data Science in all its possible applications, including small and large problem exploration and solving.

Example: A faculty member with an idea for a project could schedule an appointment at the Center to discuss their idea and learn about the pros and cons of various programming languages and which one they may want to pursue to move forward with their idea.
Example: A graduate student could speak with someone at the Center about pros and cons of different statistical approaches for solving a problem (e.g. If they have data for which they wish to develop a classification rule, should they consider a decision tree approach, a logistic regression approach, a neural network approach, etc).

Example: A faculty member is planning a course, and could consult with the Center to help find available datasets that lend themselves to relevant examples of the types of data analysis skills their students are learning.

The library, as a natural home for curating and disseminating knowledge, will play a key role in this aspect of the Center. Trends and issues in our technological and educational environments inevitably impact and affect academic libraries in higher education. The top 2018 trends and issues share several overarching themes including the impact of market forces, technology and the political environment and libraries. Specific trends and issues regarding the publisher and vendor landscape; OER (open educational resources); research datasets acquisition, text mining, and data science; and collection management, e.g. open access collection development policies and funding schemes, are relevant to the launch and sustainability of Applied Data Science initiatives at Seattle University.

The launch and sustainability of Applied Data Science at Seattle University presents numerous opportunities for Lemieux Library Faculty and staff to work with the Center and ITS to draw on, customize and extend best practices, as illustrated below.

With the growth of data science and quantitative research needs, academic library collection managers have engaged in the establishment of more defined guidelines and best practices for the acquisition of standalone spatial and quantitative datasets. Data sources now include text and numeric data, multimedia data, social media data, and hypertext and hypermedia data. Relevant mining techniques and methods range from information extraction, information retrieval, natural language processing, classification, and clustering to different ways of text summarization.

Datasets possess their own sets of acquisition and management challenges, including licensing restrictions, access and ownership, support, maintenance, discovery, and cost. Some academic libraries are beginning to offer more secure and dedicated funding lines for research datasets. Lemieux Library will work with the Center to determine the best means of managing, funding, and developing these small data set collections.

Data sources are usually in silos and use different standards, rendering data integration difficult, posing challenges to academic librarians and researchers alike. Datasets containing sensitive information, such as social media data, enterprise data, and health data pose additional challenges. The Center will need to develop privacy-preserving techniques to be applied carefully throughout the data integration, sharing, and processing stages. Both the Lemieux Library and Seattle U’s Institutional Review Board have relevant expertise that can be leveraged to this end.

The Center can assist researchers by facilitating access to data. Getting access to data remains a significant challenge. Many datasets are copyright-protected, and fair use rights could be limited by licenses. There are still a variety of approaches among vendors for access to their respective corpus of data/text. These approaches may or may not be in line with academic
library best practices or library technical capabilities (e.g., dedicated servers for storage or development of content requirement of local developer resources to support).

The Center can also assist researchers by clarifying legal aspects and negotiating licensing permissions with publishers. By creating guides on text and data mining tools and methods and providing information on databases and data sources, the Center can support training and awareness of the data resources and tools that they purchase. The Center can also provide support in areas such as digitization, data extraction, data preparation, and devising models for data analysis. At the end of text and data-mining projects, the Center can work with the Lemieux Library to preserve the datasets for reuse, assist researchers to contribute to open access datasets, and record metadata for discovery.

The Center can also be a thought leader on best practices for the ethical use of data sets and analytics. Disruptive technologies and other innovations are creating businesses that gather and use data in new and unforeseen ways. Multichannel marketing, ecommerce, fraud prevention, health care research, social media, new financial platforms, cryptocurrency systems, and a variety of other products for direct use by individuals are architected to create massive consumer data sets. However, the businesses gathering, analyzing and using these data sets often are unbounded by any form of usage constraints other than their own ingenuity and financial considerations. In addition, many of these data sets only become available through lax security practices or data security breaches. Just because data sets are available does not mean they should be used and analyzed. The Center could be a pioneer in examining these usage scenarios and developing guidelines on the ethical use of data for scholarship, teaching, and business. The establishment of data science programs at numerous institutions has led to the need for librarians to adapt and integrate growing management, accessibility, and technical subject expertise to support data scientists. As a vital part of the Applied Data Science Center, Lemieux Library Faculty will need to develop the knowledge, skills and abilities necessary to help students, faculty and staff use and address big data questions.

Given that open access, open science, open data and data science are increasingly inextricably linked, discussion of research datasets acquisition, text mining, and data science would be incomplete without consideration of open access. It will be incumbent on the Center and the Library to establish clear policies that outline parameters for the support and funding of specific open access initiatives and programs. In addition, Library Faculty will need to discuss how to incorporate open access developments into collection decision-making in relation to ever-increasing serial budgets.

This coordinated effort by the Center, ITS, and the Library will help streamline the ability of students and faculty to accomplish the following:

- Accessing public and private datasets
- Pulling, merging, and analyzing data
- Looking for patterns or trends
- Using a wide variety of tools, including R, Tableau, Python, Matlab, Hive, Impala, PySpark, Excel, Hadoop, SQL and/or SAS
- Developing and testing new algorithms
- Simplifying data problems
- Developing predictive models
- Building data visualizations
- Pulling together proofs of concepts
• Writing up results to share with others
• And other data science-related activities

3. **Staffing**

Ideally, the Center should be staffed and facilitated by a full-time Director supported by affiliated faculty from all disciplines across the University and advisory boards that include alumni and other appropriate people. A full-time Director would optimize continuity from term-to-term and allow the University to have a point-person to keep focus on the various applied data initiatives being proposed. However, in a fiscally challenged environment, the Center might be led by a faculty director or co-directors serving on a part-time basis; and use graduate and/or undergraduate student assistants as opposed to dedicated staff.

In addition, the increased focus on Applied Data Science surfaces the potential need for a dedicated Data Science Librarian to be added to the Lemieux Library staff. This need should be assessed as a part of the staffing decisions.

4. **Applied Data Science Event Coordination**

The Center should be resourced and authorized to establish and conduct: an Annual National Applied Data Science Conference; monthly or quarterly seminars and colloquia; annual “boot camp” for University-wide data literacy; and a semi-annual “hack-a-thon.”

*Example:* Each year the staff leader for the Center through a to-be-determined process would choose an emerging technology, such as blockchain, and perform a deep dive into the technology. At the end of the year, they would present their findings via Seattle University publications (Alumni newsletters, etc.) and at a speaking event.

*Example:* Throughout the year the Center would collect ideas for analytical projects. Annually or semi-annually, the student center would be reserved for a weekend long hack-a-thon where groups of students would work in teams for long hours on those projects. At the end of the weekend they would present their progress to a panel of judges made up of industry professionals and faculty. To encourage diverse program involvement, projects would be scored on not only technical and math merits but also on business pitch, legal concerns, ethical concerns, social impact, etc.

**Current State**

It should be noted that in the absence of a formalized University Applied Data Science Center that provides scientifically optimal and financially smart infrastructure, faculty and students are making partial use of cloud resources as best they can. However, this has led to additional complexity and costs for individual students, some of whom are paying for cloud platforms for group projects on personal credit cards. This current approach is not scalable or sustainable and creates barriers to scholarly success for students and faculty. We expect increased faculty and student demand for appropriate infrastructure, capabilities and tools for their use.

The following table provides specifics regarding some of the technologies and/or capabilities required to pursue Applied Data Science projects, along with comments regarding Seattle University’s current state.
<table>
<thead>
<tr>
<th>Technology / Capability</th>
<th>Seattle University Current State</th>
<th>Actions to Achieve Desired State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
<td>Ubiquitous wired and wireless connectivity, along with high bandwidth internet connectivity for all campus users.</td>
<td>Add simplified guest access to SU network to foster improved collaboration</td>
</tr>
<tr>
<td>General office productivity tools</td>
<td>Microsoft Office 365 available for all faculty, students and staff.</td>
<td>Add resources for end user training and real-time support.</td>
</tr>
<tr>
<td>Learning Management</td>
<td>Canvas LMS is the Seattle University standard for all but the Law School.</td>
<td>Continue.</td>
</tr>
<tr>
<td>General Collaboration</td>
<td>No single standard. Various products available through Seattle University license (e.g. SharePoint, OneDrive, personal drive, BePress – scholarworks.seattleu.edu), or via other external services (e.g. Dropbox, Slack, Google Hangouts, Overleaf).</td>
<td>Establish recommended approaches and deliver fully supported collaboration platform(s) and services.</td>
</tr>
<tr>
<td>Available computing capacity/servers</td>
<td>Varies by program and ability of each individual faculty member. Some programs have robust on-premise computing capacity (e.g. CS), and some faculty members leverage Amazon AWS or other cloud services for their research or class projects. Many programs and faculty members do not have the computing capacity they need.</td>
<td>Deliver standard approach to leverage cloud services.</td>
</tr>
<tr>
<td>Virtualization / Containerization</td>
<td>VMWare is extensively used by IT Services and some academic programs. Varied approaches used by other departments. No university-wide licensing for relevant products (e.g. Docker).</td>
<td>Provide guidance regarding recommended methods.</td>
</tr>
<tr>
<td>Storage</td>
<td>ITS provides traditional storage options (file folder-based) to academic departments, and certain departments augment this with additional capabilities they fund and manage, both on premise and in the cloud.</td>
<td>Expand access to cloud-based storage options.</td>
</tr>
<tr>
<td>Available Datasets</td>
<td>Various datasets available from different sources for free/for fee.</td>
<td>Provide simplified methods to identify and leverage datasets.</td>
</tr>
<tr>
<td>Extract / Transform / Load tools (ETL)</td>
<td>No standard/preferred/recommended tools for academic programs - varied approaches.</td>
<td>Procure standard tools aligned with preferred cloud services and storage.</td>
</tr>
<tr>
<td>Big Data Manipulation</td>
<td>No standard/preferred/recommended tools for academic programs - varied approaches.</td>
<td>Recommend and/or procure standard tools and services with flexibility for varied approaches.</td>
</tr>
<tr>
<td>Technology / Capability</td>
<td>Seattle University Current State</td>
<td>Actions to Achieve Desired State</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Analytical tools</td>
<td>Generally, individual departments have purchased academic program-specific products. This limits the availability of tools/capabilities that might be useful across the institution (e.g. MATLAB, Atlas.ti).</td>
<td>Recommend and/or procure standard tools and services with flexibility for varied approaches.</td>
</tr>
<tr>
<td>Data visualization</td>
<td>No standard/preferred/recommended tools for academic programs - varied approaches.</td>
<td>Recommend and/or procure standard tools and services with flexibility for varied approaches.</td>
</tr>
<tr>
<td>Virtual / Augmented reality</td>
<td>No standard/preferred/recommended tools for academic programs - varied approaches.</td>
<td>Continue</td>
</tr>
<tr>
<td>Programming Languages</td>
<td>No standard/preferred/recommended programming / scripting language for academic programs - varied approaches.</td>
<td>Recommend and/or procure standard tools and languages with flexibility for varied approaches.</td>
</tr>
<tr>
<td>Code sharing and collaboration</td>
<td>Most groups use freely available platforms – GitHub is the most popular, but there are others, as well.</td>
<td>Recommend standard tools and services with flexibility for varied approaches.</td>
</tr>
<tr>
<td>Procurement and Billing</td>
<td>Most traditional software and all Seattle University-wide software licenses and common services are purchased through standard procurement in ProcureSU. Many data science-related products and services are now procured through the cloud on a department credit card if consumption extends beyond free vendor credits.</td>
<td>Implement standard billing methods for standard tools, platforms and services.</td>
</tr>
<tr>
<td>End user training</td>
<td>There is no central source for training or a repository for knowledge-sharing for those who pursue Applied Data Science projects - everyone is just doing the best that they can.</td>
<td>Provide resources and a knowledge-sharing platform through the Center.</td>
</tr>
</tbody>
</table>

Because of the above, most projects must make decisions about the products and services they will use for every step of the work, with a few strong default options available. Some of this variation is useful based on the unique needs of a class or research project, but much of this variation is not, and ultimately limits cross-team and cross-program collaboration.

**Conclusions and Recommended Actions**

Given the rapidly changing landscape of cloud services and available technologies for applied data science, reasonable freedom is necessary for faculty and students to select and utilize varied on-demand services for classroom work and research. Allowing for this, the process of the Center to provision infrastructure will be iterative to provide the greatest benefit to the faculty and student practitioners at the most reasonable cost. The Center should be established with governance defined in such a way to not unreasonably restrict experimentation with new cloud
services and other enabling technologies not yet fully integrated into the Seattle University infrastructure. Though impossible to fully curate all aspects of cloud services for campus use, we recommend designating standard or at least preferred vendors, products and services wherever possible to provide comprehensive access for faculty and students, streamline end-user start-up, coordinate support services, provide training, and manage costs and risks.

Implementing institutional agreements with preferred vendors will provide favorable terms to Seattle University and reduce risk and costs when compared to the standard end user license agreements (EULAs). Implementing supporting technical infrastructure for preferred vendors, products, and services will simplify access and usage by end users. Specific technical infrastructure may include predefined network connectivity and security, identity and access management, license distribution and management, usage reporting and standardized billing, or templates and predefined automation.

We recommend the following next steps:

**Recommendation #1:** Commit to establishing a Seattle University Applied Data Science Center. Provide resources to the center from internal sources as well as from external sources such as industry, foundations, and individual donors.

**Recommendation #2:** Establish the Center as a hub for all data science activities. This would include:

- Identify the initial collaboration platform(s) to be promoted by the Center for use by faculty, students, and staff in staying abreast of currently available capabilities.
- Reach out to faculty with existing teaching and research obligations to assure that needs are properly assessed and tracked.
- Provide general guidance to the Seattle University community regarding the use of available Applied Data Science vendor/product/service capabilities. Share how input for changes and improvements will occur.
- Establish Applied Data Science governance relationships and processes. Governance would include: decision-making regarding desired products, services and tools for Applied Data Science; formal requests for prioritized infrastructure investments; policy recommendations; intellectual property guidance; and other relevant decisions. Governance partners would include the Director of the Applied Data Science Center, the Faculty Technology Committee, the Chief Information Officer, and others.

**Recommendation #3:** Ensure adequate infrastructure and resources for data science initiatives. This would include:

- Establish an annual budget for Applied Data Science infrastructure investments and identify an appropriate source of funding.
- Hire a technical resource within IT Services (1 FTE) to focus on cloud infrastructure implementation for Seattle University. This resource would implement enabling technologies (e.g. network, identity, security) and partner with the practitioners/coaches in the Applied Data Science Center and the various academic programs.
- Investigate and assess the need for an Applied Data Science Librarian.
- Create and maintain a multi-year rolling roadmap of planned Applied Data Science infrastructure investments (includes back-office technologies, end-user tools/product/services purchases) and deliver these improvements per the master schedule.
Using the center, create a strategy to garner resources and support from industry, foundations, and individual donors for applied data science initiatives campus wide.

Conclusion

Today we know that data science is rapidly becoming woven into the fabric of organizations of all sizes and types. Further, we know that data science is driving significant societal and economic impact. Organizations of all types are increasingly becoming data driven, investing in infrastructure, people and processes to embrace the data science journey.

The recommendations in this report focus on three areas – the data science landscape, curricular and programmatic initiatives, and an Applied Data Science Center. Comprising an exciting vision for applied data science at Seattle University, these recommendations provide a strategy board for Seattle University to develop a timeline, budget and allocate resources on various activities along the data science process.

Investment in applied data science at Seattle University promises to serve several purposes. Most importantly, it opens up a world of opportunities for Seattle University undergraduates and graduates alike. Investment in applied data science ensures that Seattle University will continue to transform our students’ college-to-career competitive performance.

Investment in applied data science also promises to enhance our University brand and accelerate our growth. Seattle University is known for its values-infused quality connections and close collaborations on and off campus, offering our community, educational, institutional and corporate partners involvement in projects that span across disciplines, departments, schools and colleges. This presents a distinct competitive advantage in a world where the solutions to many global challenges lie in the intersection of research, innovation, technology, and business. Within the many possibilities of connecting discipline, department, school and college, applied data science promises to elevate students and our partners alike.