Library Partnerships to Support Data Analytics Engineering Programs

Ms. Wendy Mann, George Mason University

Wendy Mann is the Director of the Digital Scholarship Center (DiSC) at George Mason University Libraries.

Ms. Theresa M. Calcagno, George Mason University

Theresa Calcagno is the Librarian to the Volgenau School of Engineering at George Mason University in Fairfax, VA. Prior to that, Calcagno was a Research Librarian for an engineering project consulting firm.

Deborah Ann Kermer, George Mason University

Deborah Ann Kermer is the Data Services Research Consultant at the Digitial Scholarship Center in the George Mason University Libraries. She assists faculty and students with quantitative and qualitative data, methods, and software.

Library Partnerships to Support Data Analytics Engineering Programs

Abstract

In the last decade, the number of graduate programs in Data Analytics has grown exponentially. Academic libraries have had to, or will need to, determine how to support the growing student population in this new area of academia, as well as acquire new resources and develop new services and tools for analytics students. Data Analytics and Data Science programs are particularly challenging to support because they are trans-disciplinary—incorporating areas such as statistics, computer science and business.

At George Mason University (Mason) the Data Analytics Engineering (DAEN) program is a master's degree program in the Volgenau School of Engineering. This program started in 2014 and enrolls students in 10 different concentrations including business analytics, cyber security analytics, financial engineering, predictive analytics and more. This paper discusses how the Engineering Librarian collaborated with the Mason Libraries' Digital Scholarship Center (DiSC) to identify and develop ways to help meet the needs of this unique group of students.

Introduction

Over the past seven years, master's programs in data analytics have grown quickly, as demonstrated by the steeply upward trend in enrollment reported by the Institute for Advanced Analytics [1]. The incentive for universities to add these programs, and for students to enroll, is fueled by the increasing demand for workers with skills in this area. In 2017, research by IBM Analytics, the Business-Higher Education Forum, and Burning Glass Technologies forecast that by 2020, the number of Data Science and Analytics (DSA) job openings would increase by 20% (364,000) over the number of openings in 2015 [2].

To meet the increasing demand for skilled DSA graduates, U.S. universities and colleges have responded by developing and offering degree programs in this area. From the beginning of 2014 through 2018, the total number of programs offering master's degrees in data science and analytics increased nearly six-fold from 35 to 241. By the end of 2018, the cumulative number of master's degrees awarded in these programs was approaching 20,000 [3].

When new academic programs are created, libraries must determine how to assess faculty and student needs and provide support. In addition, libraries may need to acquire new resources and develop new services and tools specifically for those students. Determining how to meet the needs of DSA is not an easy task; supporting these programs can be particularly challenging because of their trans-disciplinary nature and the varied subject content of the required courses. To be successful working with a DSA program, a library needs to ask questions of faculty and program directors to try to determine expected and potential student needs. As will be addressed, our experience is that these students' needs push the boundaries of traditional academic library

support. However, the more planning done prior to the launch of the program, the better for all parties involved.

Our purpose is to provide insight on how the librarians and professional staff in one university library collaborated with one another and program faculty to identify the needs of students in the University's DSA graduate program, and then collectively developed and implemented support services and tools to address those needs.

Background

About the University

George Mason University is a comprehensive, doctoral institution which is classified a Research 1 institution [4]. Located in the Washington, D.C. metropolitan area, the University comprises 10 colleges and schools with an enrollment of over 37,300 graduate and undergraduate students in Fall Semester 2018 [5]. The Volgenau School of Engineering offers programs in traditional engineering disciplines, computer science, cybersecurity engineering, data analytics engineering, IT and statistics.

About the University Libraries

The George Mason University Libraries encompass four libraries on three campuses. The main research library is the largest and it provides research services and collections in all subject areas. The libraries on other campuses serve specific academic programs and research centers located on those campuses. The University Libraries provide a variety of services to support instructional faculty and students. The Engineering Librarian acts as a liaison between the University Libraries and Engineering programs to make them aware of new library services for researchers and students. The Librarian also provides research, instructional, and collections support to all programs in the School of Engineering.

The George Mason University Libraries Data Services Group was established in 2011 to meet the data use and acquisition needs of the university community. Along with three full-time staff, the unit managed a computer lab. In 2017, the group's name was changed to the Digital Scholarship Center (DiSC) when its role was expanded to include additional modes of digital research, such as digital humanities.

In Fall Semester 2014, the Volgenau School of Engineering at Mason launched a graduate program in Data Analytics Engineering (DAEN). The program is trans-disciplinary in nature, leveraging the expertise of faculty across multiple University schools and departments. This cooperation was necessary to develop the curriculum and teach the program's classes. Originally, classes were only offered face-to-face and scheduled at night to meet the needs of working adults in the metropolitan area. In Spring 2019, the University, in partnership with [online learning partner] [6], expanded DAEN's reach by introducing an online MS degree that can be completed in 18 months [7]. Students in the program receive an MS degree in DAEN and may elect to complete a specific concentration chosen from applied, business, cyber, health data, predictive or

statistical analytics as well as bioengineering, data mining, financial engineering, or Internet of Things [8]. A graduate certificate is also an option.

In addition to the traditional master's program, qualified undergraduates can enroll in an accelerated DAEN master's program. Students majoring in Applied Computer Science, Bioengineering, Computer Science, or Systems Engineering can receive a BS degree in their major and an MS degree in Data Analytics in only five years [9].

About Data Analytics

Data analytics is not yet a well-define academic discipline and not only the course requirements, but the titles of the degrees awarded vary greatly. Analysis of data from the Institute for Advanced Analytics as of December 2018 [3] shows that the plurality of degrees offered were "Business Analytics" followed by "Data Science," "Data Analytics," and simply "Analytics." However, over a quarter of degree titles were used for four or fewer programs, with 43 being unique. Degrees in that "Other" category either combined two of the above titles, included the word Applied, or mentioned specific skills like "Business Analytics and Risk Management" or "Data Mining and Predictive Analytics."

Table 1. Data Science and Analytics Programs by Degree Title (as of December 2018) [3]

	Programs Offered	
Degree Title	#	%
Business Analytics	86	34 %
Data Science	58	23 %
Data Analytics	23	9 %
Analytics	20	8 %
Other	66	26 %
Total	253	100 %

Data analytics programs are found in many different types of institutions and the focus of each program varies accordingly. Of the 40 programs offered by technically-focused schools or colleges (Computing, Information, Engineering, Technology), 30 were a form of Data Science (21 exactly, 9 with a close variant). Only 4 degrees included the term Engineering and all are offered by Schools of Engineering. Whereas, of the 119 programs offered by a Business School (47% overall), over three-quarters (n=92) were labeled as Business Analytics (83 exactly, 9 with a close variant), and only two as Data Science. About a quarter of programs (n=66) were offered by a general school, such as Arts and Sciences or Graduate and Professional Education. Nineteen programs were offered jointly through multiple schools/colleges or at an institute, and 9 fell into the Other category (e.g., schools of medicine or public policy) [3].

The Data Analytics Engineering program at Mason is “aimed at students who wish to become data scientists and analysts [10].” Graduate Assistants (GAs) who were in the DAEN master’s program have considered it to be a “Data Science” program. Indeed, much the same way that an engineer applies math and science to solve problems, a Data Scientist applies statistics and other data analysis techniques to solve problems.

A Data Scientist uses more sophisticated computational tools than a Data or Business Analyst. Further, Data Scientists are typically expected to find patterns rather than answer a specific stated question. Although the goal is still to solve or address a real-world problem (e.g., best predict an outcome), it is easy for students to overlook that. We have found that students doing their capstone project are often hoping to obtain data that best shows off their skills, without regard to data quality or even meaning, much less topic.

Students enrolled in the DAEN program have diverse academic backgrounds and experiences. Some students enter the program with degrees in computer science, IT, or engineering. These students bring coding skills and computational adeptness, but often lack the well-rounded background one gets from traditional liberal arts fields of study. In contrast, other students arrive with domain knowledge based on degrees in their specific discipline, but often need to develop the analytical, computational and coding skills required to be successful in the program. Each group – the IT group and the domain-knowledge group – needs to develop a new skill-set.

Providing appropriate support and tools to students with such varying skillsets and knowledge is challenging. The rapid growth of the DAEN program at Mason resulted in a large demand for support from the Library’s Digital Scholarship Center (DiSC) and the Engineering Librarian. Out of necessity, the librarians and professional staff began working with one another to respond quickly to meet the needs of the DAEN students. Eventually, we invited faculty to contribute to our collaborative efforts for us to learn more about student and faculty needs. We also shared information about DiSC and discussed how the experts there could help both groups.

Collaborating to Meet the Needs of DAEN Students

Collaborations are common in the academic library community and range from partnerships involving multiple libraries to micro-collaborations between two subject librarians or between a librarian and a faculty member [11]. Atkinson (2018) provides an overview of different types of collaborations involving academic libraries and librarians [12]. He identifies several main types including internal collaboration, collaboration with faculty, collaboration with other support departments (e.g. writing centers), and collaboration with students [12]. In one collaboration [13], faculty and librarians worked with students in a large environmental science class who were from a wide variety of majors. Collaborators found that the students’ technical and library research skills varied widely, creating challenges [13]. We have experienced the same with DAEN students. The lack of library skills, especially those related to finding and using datasets, is compounded by the student’s need to learn new analytic techniques.

Ideally, a collaboration starts before the associated project or program is launched. However, the collaboration between the George Mason University Libraries and DAEN faculty was not pre-planned--it started at the point of need and has developed organically over the years. This type of

development can be successful if the partners involved are flexible, action oriented, and not afraid to fail. The nature of our collaboration has allowed the George Mason Library to identify current needs, pilot solutions, and revise them quickly as needed. The ability to adapt to frequent change is optimal in working with a quickly evolving program.

There can be drawbacks to this type of collaboration, a major one being that the actions taken are more likely to be reactive rather than proactive. This reactionary approach results in situations where library staff are continually working to catch up with students' changing needs. Another drawback is that Library staff contact with DAEN faculty was limited in the beginning, and somewhat serendipitous. Information collected in pre-planning could have made the identification of appropriate faculty partners and potential issues much easier.

Ivey [14] used Schrage's list of factors for successful collaborations [15] when interviewing faculty and librarians to identify behaviors necessary for the success of a collaboration. Behaviors that the groups considered necessary for success were: A shared, understood goal; mutual respect, tolerance and trust; competence for the task at hand; and ongoing communication. For a faculty-librarian collaboration to be successful, a shared goal or vision is required. If partners do not understand each other's perspective, the partnership may be less effective. Fortunately, faculty in the DAEN program have been receptive to our involvement and have accepted our offers of help for their students. Through periodic meetings with DAEN faculty, we have developed the necessary shared vision, as well as a better understanding of the needs of the students. Building good relationships with faculty requires good planning and strong, ongoing communication with them [16]; we have been able to accomplish this even though the conversations did not start when the program did.

Helping DAEN Students

From the launch of the DAEN program in 2014 through late 2015 there were no recorded contacts with DAEN students. The first interactions with DAEN students in November 2015 were unexpected because DiSC staff were not aware of this new major. Put another way, DiSC did not know about the DAEN program until students started coming to its lab. At this point DiSC contacted the Engineering Librarian to alert her to this new development.

Over five years, the number of enrolled DAEN Master's students at Mason rose by 900 percent, as shown in Table 2 [17]. The data are based on fall official census student counts reported by Mason's Institutional Research and Reporting (IRR). The most popular concentrations are Predictive Analytics and Applied Analytics, respectively, but DAEN students are not required to specify a concentration. For the 2017-2018 academic year, most DAEN degrees conferred had no specific concentration (N=91) [18].

Table 2: DAEN Fall Enrollment Numbers by Concentration [17]

Concentration	2014	2015	2016	2017	2018	Total
Not Specified	12	40	98	142	237	529
Predictive Analytics	8	33	52	46	45	184
Applied Analytics	14	35	46	34	33	162
Business Analytics	-	-	3	26	30	59
Data Mining	3	8	16	19	10	56
Statistics for Analytics	1	8	9	12	4	34
Financial Engineering	-	-	4	7	10	21
Digital Forensics	-	1	8	6	2	17
Health Data Analytics	-	-	-	-	6	6
Statistical Analytics	-	-	-	1	4	5
Bioengineering	-	-	-	1	-	1
Total	**38**	**125**	**236**	**294**	**381**	**1,074**

Table 3 shows the usage data from DAEN users of DiSC's Data Services. In retrospect, it was surmised that some of the students self-reported their affiliation with the Volgenau School of Engineering, rather than DAEN. Although DiSC works with students in other Engineering concentrations, most of its users from that school are DAEN students. The data below include only students identifying specifically as DAEN. Clearly, there is a large increase from the 2015-2016 to the 2016-2017 academic years.

Table 3. Number DAEN Students Use of DiSC by Academic Year (July 1-June 30)

	2015-2016	2016-2017	2017-2018
Recorded Consultations	6	80	95

In June 2016, the director of DiSC and the Engineering Librarian met with the DAEN Program Director to discuss the program and student needs, especially related to finding and accessing data. This meeting helped us gain a better understanding of the program beyond what was published on its website, and we were made aware of their plans. Around this time, just as DAEN enrollment was increasing, DiSC witnessed an uptick in consultations with and lab visits from students. Not surprisingly, this trend was in line with DAEN enrollment growth.

We observed that DAEN students often had a different approach to data needs and data questions than students and researchers in other areas of study. Up to this point, DiSC data consultations were predominantly with researchers from the social and health sciences, or scientific disciplines such as geography or environmental studies. Researchers in those areas typically needed data to solve a research question and/or support a hypothesis. Those research questions are what library-based data services—such as data, social science, business, or government information librarians—are used to answering. However, many of the DAEN requests fall outside of this

realm; those students want datasets to practice on, or to build a model or a digital tool with. Thus, instead of requesting data about a topic, they ask for data with features such as: 100,000 or more observations, specific variable types (e.g., interval/ratio or nominal), or text corpora.

Examples of DAEN Data Projects:

- Create a tool to recommend places to live in the U.S. based on weather, terrain and population density.
    - For this project, the students needed average precipitation and temperature data by zip code. Three librarians from DiSC worked with this group to teach them how to identify and extract data from NOAA. There were other geographic aspects of this project that required the use of GIS software to filter by zip code. At this point, DAEN students were not learning about GIS.
- Use of American Community Survey (ACS) API to extract data to build a model.
    - In this case the group was concerned that there was something wrong with their data extraction because the data were not being continually updated. We explained how ACS data are collected and released (using a survey and released on specific dates). Further, they could simply download the data needed for the project rather than building an API request.
- Repeated requests for private, company data.
    - This group needed data so that they could demonstrate their ability to work with business data since their end goal was to find a position as a data engineer for a corporation. The type of data they wanted — about internal operations of a corporation — is typically private. Issues with proprietary and private data and that not all data is publicly available was explained to them. They eventually modified their topic to work with available company data.

In the above examples, each group encompassed students with technical backgrounds. They had the computer skills to build models and write code, but they lacked understanding of the basics of working with data or the data's subject area (e.g., deciphering documentation and understanding how surveys work). The latter concepts are taught to researchers in data-intensive disciplines such as sociology, psychology, public policy or public health, but did not appear to be addressed in the DAEN curriculum. Further, for their group project topics, students were looking for data to show off their computational skills, rather than to solve a problem. In talking with the students, we learned that many with technical backgrounds were used to working with machine data and were new to using "messy" human data.

Solutions

To date, DiSC has employed three graduate assistants (GA) who were DAEN majors. They were quite helpful in working with us to fulfill these new DAEN student needs. Fortunately, only two years after the DAEN program launched, one of the DiSC GAs transferred into the DAEN program. Her support was invaluable for clueing us in to what to expect, such as class assignments and what data sources we could make available. As available, we reviewed syllabi and degree requirements to gather more information.

DiSC GAs have also assisted with and led workshops and consultations. Additionally, four other GAs were from other computational or data intensive majors: computer science, biostatistics, and psychology. Relying on peers to assist with consultations for the DAEN student has worked well. In the peer-to-peer consultations, DiSC staff feels that the DAEN user is more likely to accept what they are being told regarding limits for data access. One problem, though, is that the GA consultants need to be constantly reminded not to provide too much assistance in coding to avoid doing the student's assignment. This is not an issue with other departments because computer science programs have strict guidelines about accepting help, and coding is not an assessed skill in others and thus help can be freely given. But, when a DAEN student is struggling, GAs must be instructed to talk to full-time DiSC staff, which often results in the student being referred back to their professors.

All of DiSC's DAEN GAs had undergraduate degrees outside of engineering and computer science. They had well-rounded experience working with data, which provided a broader perspective with their consultations. Full-time DiSC staff also consulted with them when assisting DAEN students. The GAs recommended data sources used by students in DAEN coursework, which helped us to build a subject guide page of data for "Practice & Projects" that specifically addressed DAEN data requests and those from related majors. Below is a screenshot of the page.

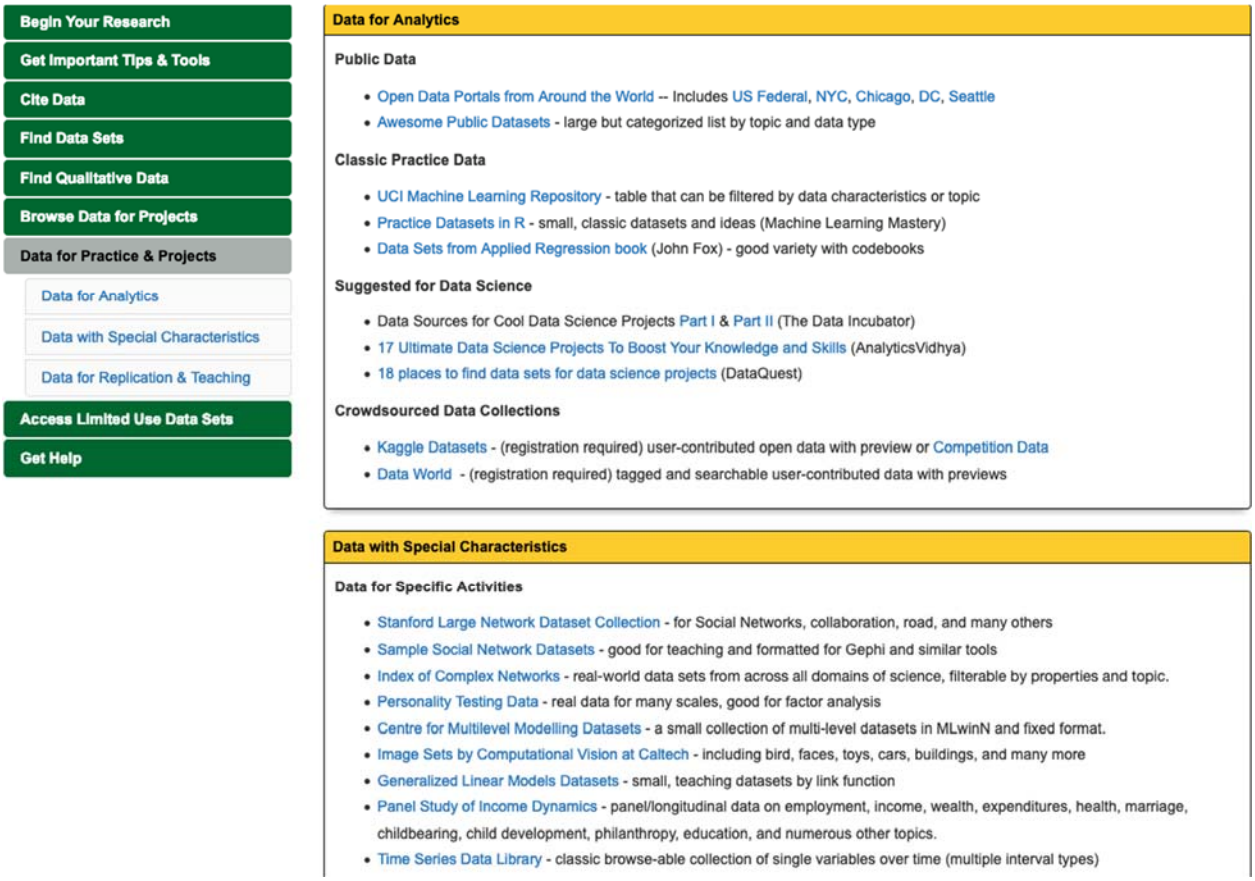

Figure 1: LibGuide page designed for DAEN data requests.
Source: https://infoguides.gmu.edu/find-data/practice

DiSC staff, along with the Engineering Librarian, continue to reach out to faculty from the various departments that teach required DAEN classes. We have made connections with the Capstone instructors, Program Director, and some DAEN-affiliated faculty in other departments. The program director has recently created a "Foundations" class that "may be required for students without a basic foundation in Data Analytics" [19] to address gaps in students' knowledge. Going forward, we hope to be able to contact the students early in the program through that class. In addition, Capstone project groups are being instructed to consult with the Engineering Librarian and DiSC staff early in the semester to ensure that data for their projects are available and accessible.

It is important to educate all students, not just those in DAEN, about concepts such as copyright, licenses and terms of use. For instance, raising student awareness regarding licensing issues for projects using a proprietary database such as Nexis Uni® is important. Automated text and data mining are not allowed in Nexis Uni® and other library databases, thereby restricting some projects proposed by students. We need increased opportunities to teach ethical and legal issues in working with data and to teach students that not all data are free and unrestricted.

Data analytics is different from other academic programs because the learning outcomes are about developing techniques and selecting analyses; instead of instilling knowledge about a subject, the goal is to develop a skill-set. Students in traditional research domains, such as public health or economics, are also learning data analysis and many of the same skills. But, although some domain-experts do decide to enhance these skills by entering a data analytics program, the primary cohort of DAEN students do not have such a background. As one faculty member told us, DAEN students are taught that in the real world they will need to include domain experts on projects. The hope is that this idea is reinforced as they move forward on their capstone projects.

Conclusions/Lessons Learned

Aligning information literacy instruction with the needs of students and faculty researchers is essential. Teaching how to find, use, and manage data in research is increasingly important as students conduct more primary data collection or use repositories to locate secondary data for class projects [20]. To teach this, librarians must expand their own skill-sets to include digital and data tools, as well as stay on top of emerging research areas and methods. Libraries also need to support a wider variety of research needs, and consider purchasing new sources of data, such as social media data or products that allow text and data mining of current news.

We believe that the collaboration between the Libraries and the DAEN faculty has benefitted students in the program. As the cooperative effort has developed along with the DAEN program, the Libraries have been better able to quickly identify and respond to its needs. Major outcomes of this collaboration are:

1. Increased interaction with the program director and the instructors of DAEN and associated courses.
2. DiSC professionals and the Engineering Librarian developed a better understanding of DAEN students' abilities to find, work with, and ethically use data, and to conduct literature research.

3. The faculty in DAEN and associated departments are more aware of the services and tools that DiSC offers and refer students to DiSC for assistance.
4. The faculty of the Capstone program are more aware of the common difficulties that students encounter.

In the future, our plans are to:

1. Continue to develop our relationship with the DAEN Program Coordinator and other faculty in order to stay informed of changes to the program or new needs that may arise.
2. Visit the newly created 500-level DAEN "Fundamentals" course [19] to orient students to the University Libraries and introduce them to DiSC and its services.
3. Offer a pre-capstone workshop for DAEN students the semester before the Capstone class. In this workshop, library staff could discuss developing a research question and a realistic project scope, data sources and license restrictions, and more. Learning to develop a realistic project scope is especially important for students since the project must be completed in one semester.

References

[1] M. Parry, "Data Scientists in Demand: New programs train students to make honest sense of numbers," *Chron. Higher Educ.* Mar. 04, 2018. [Online]. Available: http://www.chronicle.com/article/College-Rush-to-Ride/242674?cid=cp188
[2] W. Markow, S. Braganza, B. Taski, S.M. Miller, and D. Hughes, "The Quant Crunch: How the demand for data science skills is disrupting the job market," Burning Glass Technologies, Boston, MA, USA, 2018. [Online]. Available: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IML14576USEN
[3] Institute for Advanced Analytics, "Graduate degree programs in analytics and data science." Accessed Jan. 21, 2019. [Online]. Available: https://analytics.ncsu.edu/?page_id=4184.
[4] George Mason University, "Mason achieves top research ranking from Carnegie," Feb. 3, 2016. [Online]. Available: https://www2.gmu.edu/news/182106#
[5] Office of Institutional Research and Effectiveness, George Mason University, "2018-2019 Facts and figures," Dec. 2018. [Online]. Available: https://irr2.gmu.edu/FastFacts/
[6] P. Williams, "Mason expands online graduate offerings through Wiley partnership," Jun. 28, 2018. [Online]. Available: https://www2.gmu.edu/news/513161
[7] Mason Online, "Data Analytics Engineering (MS)." Accessed Jan. 23, 2019. [Online]. https://masononline.gmu.edu/programs/master-of-science-in-data-analytics-engineering/
[8] University Catalog 2018-2019, "Data Analytics Engineering MS: Requirements." Accessed Jan 21, 2019. [Online]. Available: https://catalog.gmu.edu/colleges-schools/engineering/data-analytics-engineering-ms/#requirementstext
[9] University Catalog 2018-2019, "Data Analytics Engineering, MS: Accelerated Master's." Accessed Jan. 21, 2019. [Online]. Available: https://catalog.gmu.edu/colleges-schools/engineering/data-analytics-engineering-ms/#acceleratedmasterstext
[10] Volgenau School of Engineering, "Data Analytics Engineering, MS: Overview," Accessed Jan. 30, 2019. [Online]. Available: https://volgenau.gmu.edu/program/view/20563

[11] E.D. Cassidy and K.E. Hendrickson. “Faculty-librarian micro-level collaboration in an online graduate History course,” *J. Acad. Librarian.*, Vol 39, no. 6, pp. 458-463. Jun. 2013. DOI: https://doi.org/10.1016/j.acalib.2013.08.018
[12] J. Atkinson, “Collaboration and academic libraries: An overview and literature review,” in *Collaboration and the Academic Library: Internal and External, Local and Regional, National and International*, J. Atkinson, Ed. Kidlington, UK: Chandos Publishing, 2018, pp.11-33. DOI: https://doi.org/10.1016/B978-0-08-102084-5.00002-X
[13] N.G. Kobzina, “A faculty—librarian partnership: A unique opportunity for course integration,” *J. Lib. Admin.*, Vol. 50, no. 4, pp. 293-314, Apr. 2010. DOI: https://doi.org/10.1080/01930821003666965
[14] R. Ivey. “Information Literacy: How do librarians and academics work in partnership to deliver effective learning programs?” *Austr. Acad. & Res. Libs.*, Vol. 34, no. 2, pp.100-113, Jun. 2003. DOI: https://doi.org/10.1080/00048623.2003.10755225
[15] M. Schrage. *Shared Minds.* New York, NY, USA: Random House, 1990.
[16] J.O. Díaz and M.A. Mandernach, “Relationship building one step at a time: Case studies of successful faculty-librarian partnerships,” *portal: Libraries and the Academy*, Vol. 17, no.2, pp.273-282, Apr. 2017. DOI: https://doi.org/10.1353/pla.2017.0016
[17] Office of Institutional Research and Effectiveness, George Mason University, “Beginning of term student enrollment by program.” Accessed Jan. 21, 2019. [Online]. Available: https://irr2.gmu.edu/New/N_EnrollOff/EnrlStsProg.cfm
[18] Office of Institutional Research and Effectiveness, George Mason University, “Degrees conferred by demographic.” Accessed Jan. 21, 2019. [Online]. Available: https://irr2.gmu.edu/New/N_Degree/DegMajor.cfm
[19] University Catalog 2018-2019, “Data Analytics Engineering MS: Admissions & Policies.” Accessed Jan 21, 2019. [Online]. Available: https://catalog.gmu.edu/colleges-schools/engineering/data-analytics-engineering-ms/#admissionspoliciestext
[20] D. MacMillan, “Developing data literacy competencies to enhance faculty collaborations,” *Liber Quarterly*, Vol 24, no. 3, pp. 140-160, Jul. 2015. [Online]. DOI: http://doi.org/10.18352/lq.9868