ASEE'S VIRTUAL CONFERENCE
At Home with Engineering Education
JUNE 22 - 26, 2020 #ASEEVC
Paper ID #28741

Research data practices of aerospace engineering faculty: A qualitative study

Fred Rascoe, Library, Georgia Institute of Technology

Fred Rascoe is the Scholarly Communication Librarian and the Aerospace Engineering Librarian at the Georgia Institute of Technology Library.

Lisha Li, Georgia Institute of Technology

Lisha Li has been a Science and Engineering librarian at the Georgia Tech Library since 2005. Currently in the Campus Engagement and Scholarly Outreach department, she is the subject liaison for the School of Civil and Environmental Engineering.

Research data practices of aerospace engineering faculty: A qualitative study

Introduction

The storage, curation, and accessibility of digital research data is becoming more important for scientific and engineering researchers at academic institutions. While much has been written on the needs of research data management of scientists in the academic literature, less common are studies that look in finer detail at specific engineering disciplines. In this paper, the authors use qualitative data from interviews with research faculty of a specific engineering discipline, aerospace engineering (AE), to identify common perceptions, practices, and strategies with respect to their research data. This report also discusses how these faculty react to, adapt to, and accommodate those challenges, often in the words of the respective researchers themselves. This report then concludes with possible recommendations for academic libraries, and research university campuses with an AE presence.

Background

In 2019, librarians at the Georgia Institute of Technology (Georgia Tech) conducted interviews with faculty in the institute's School of Aerospace Engineering to examine practices of academic research faculty with regards to the data used and produced in the course of their research. This work built on a project conducted in 2017-18, when librarians at Georgia Tech joined with librarians at 10 other research institutions, coordinated by Ithaka S+R, in order to examine the broader research practices of academics in civil and environmental engineering, the findings of which are summarized in a report [1].

Established in 1885, Georgia Tech is a public research university with over 23,000 students and an $824 million in R&D expenditures. Georgia Tech's engineering programs have been consistently ranked within the top 5 in the U.S. Georgia Tech plays a leading role in the state's economic development strategy. Research is conducted for industry and government by the applied research division of Georgia Tech, various academic schools and departments, and more than 100 interdisciplinary research units.

The School of Aerospace Engineering at Georgia Tech is ranked among the top 5 aerospace engineering programs in the country. The school focuses on 6 disciplines of study and research: aerodynamics and fluid mechanics; aeroelasticity and structural dynamics; flight mechanics and controls; propulsion and combustion; structural mechanics and materials; and system design and optimization. The school has a faculty of more than 40 tenure-track professors and enrollment of more than 1,300 graduate and undergraduate students. There are nearly 20 labs and research groups housed within the school.

Selected Literature Review

Research libraries have provided data management services for over a decade. Many such services continue to provide value and to make an impact in their respective research communities. In order to improve data services for research communities, librarians have constantly sought feedback from researchers regarding their changing needs and practices. Samuel and others analyzed twenty-nine data management plans (DMP), specifically related to National Science Foundation grant proposals, to understand better, how faculty approach data management. The results help librarians build a foundation for future DMP services [2]. Coates and others examined five case studies that highlight common challenges for librarians to evaluate existing research data services in academic libraries, and suggested that evidence-based approach provides valuable information for assessing the still-emerging services [3]. Goben and Griffin's study confirmed, "Researchers are most worried about storage, sharing, and issues that revolve around longer term access to data…" Therefore, there are needs for "thoughtfully planned academic RDS [Research Data Service] that are simultaneously broad in scope and strategically focused on addressing specific local needs" [4]. Perrier and Barnes used mixed methods to determine the essential tools and services required for research data management. They used focus groups, together with a rating exercise, to provide valuable information for libraries to identify key areas and to prioritize areas for development [5].

A few quantitative studies identified differences among research domains, disciplines and different faculty ranks. For instance, Whitmire, Boock and Sutton conducted a campus-wide survey and found that "researchers are generating a wide variety of data types, and that practices vary between colleges" [6]. Barsky surveyed science and engineering researchers, and the result showed that over 40 percent of researchers in science and engineering reported insufficient documentation in research data practice [7]. Akers and Doty revealed significant differences among different ranks of faculty members in their views on research data management and data sharing habits [8].

While there have been many research studies focusing on research data practices across disciplines, a handful of studies within the last a few years adopted qualitative or mixed methods, in analyzing research data practices for science and engineering related fields. Carlson summarized the findings of a research workshop conducted by a team from the Purdue Universities Libraries and the Graduate School of Library and Information Science at the University of Illinois and described the development of the Data Curation Profile Toolkit (DCP Toolkit) that resulted from the study. This toolkit set a foundation for librarians to conduct data interviews with their researchers in an individual or group settings [9]. Using a combined approach, Van Tuyl and Michalek assessed researchers' data management practices by implementing a survey and conducting a number of interviews with a sample of faculty. They found that the types and quantities of data produced by scholars differ in each institution and among research domains [10]. Parham et al. also conducted a faculty survey at a technical institute. The survey was both quantitative and qualitative in nature, based on the Data Asset Framework (DAF) template. This survey explored research data needs of this particular campus and found that "73% of our respondents indicated an interest in data storage and preservation," where respondents came from a wide range of departments and disciplines [11]. Wiley and Mischo's study indicated, "The disciplinary differences in data management needs are significant and represent a set of challenges for libraries in setting up consistent and successful services" [12]. Later on, Wiley's study on data sharing among engineering faculty revealed, "There is no correlation between Open Access (OA) journals and research data sharing practices" [13].

In response to researchers' needs, librarians have incorporated data literacy into instructions. Holles and Schmidt described the development of a graduate course on research data management co-taught by a librarian and a faculty member while exploring new tools such as cloud computing, data visualization and simulation, data analytics and others [14]. Built upon the result of interviews of science and engineering faculty, Mischo and others developed instruction and support programs that integrate data management support into scholarly communication instruction and training, and enhanced it with contents aimed at specific discipline and research practice variations [15]. From the library and information education perspective, Cox and others indicated that academic libraries have turned their focus to "a greater emphasis on research and with a concomitant greater need to engage with issues around data such as data integrity and data analysis" [16].

Few qualitative studies examined data practices in one specific discipline. One recent study is the Ithaka S&R's research report, as mentioned earlier [1] that examined the changing research methods and practices of civil and environmental engineering scholars based on research studies from 11 academic libraries in the U.S. and Canada. The report, based on the result of in-depth qualitative analysis, highlighted the findings of how civil and environmental engineering scholars discover, manage and share research data as a part of the integral study. It clearly showed that there are significant inefficiencies in current data sharing practices within civil and environmental engineering scholars.

So far, the authors have not seen literature on qualitative studies of research data practices focusing on AE specifically. Built upon previous studies, the authors see a need to conduct a qualitative research of AE scholars and to better understand how they work with research data, and what challenges they are facing. The authors anticipated that the result of such a study could shed light on how to improve, expand or establish robust, discipline specific data management services and data education programs on the campus.

Methodology

The methodology of this qualitative study was adapted from the 2019 Ithaka S+R Civil Engineering Project mentioned above [1]. The authors submitted a study proposal for this project to the local Institute Review Board. After approval, the authors met with the chair of Georgia Tech School of Aerospace Engineering to explain the purpose of the project, gain administrative support, and to help identify faculty members with wide ranging research interests and practices. These faculty members were then recruited via email (see Appendix 1). Nine AE faculty members, representing a wide range of areas of research and practice, committed to the project. Of the AE faculty members that participated, the primary areas of study included Aerodynamics & Fluid Mechanics (5 faculty), Systems Design & Optimization (2 faculty), Flight Mechanics & Controls (2 faculty), Aeroelasticity & Structural Dynamics (1 faculty), and Propulsion &

Combustion (4 faculty). Most faculty interviewed have at least two listed primary areas of study, which accounts for the numbers being greater than nine.

Individual, semi-structured interviews with these AE faculty were conducted and recorded over a three month period. The interview questions focused on the following topics: faculty members research areas, collaboration, research data, and future challenges to the field. Faculty were given a consent form to agree to verbally (Appendix 2), and were given the interview questions in advance (Appendix 3). Interviews averaged 30 minutes in duration, and the digital recordings were transcribed by a professional transcription service. These transcripts were then reviewed by the authors for quality control. The authors anonymized all transcripts.

The approach to coding was based on a grounded approach for qualitative analysis of data, as defined by Gordon and Strauss [17], and used in the coding approach adopted in the 2019 Ithaka S+R Engineering project. The authors reviewed anonymized transcripts to do an open coding analysis. Next, the codes were grouped into broader categories or themes. The authors discussed points of overlap found via coding analysis, as well as unexpected or unanticipated points of interest raised by the interviewees. These results of this analysis are presented here.

Results

Research Methods and Collaboration

Broadly, faculty in the School of Aerospace Engineering can be divided into those whose research is experimental, and those whose research is theoretical. The experimental work among respondents is largely centered in the important aerospace processes of combustion, propulsion, aerodynamics, and dynamics and control. Experimental work often takes place in laboratory spaces on campus dedicated to aerospace research.

Theoretical work involves computational modeling and simulation of these same processes of combustion, propulsion, and aerodynamics, as well as aircraft design simulation, computational fluid dynamics modeling, and computational elasticity. In contrast to the experimental researchers, theoretical work makes use of high-performance computing resources, whether hosted in the lab, or hosted by Georgia Tech's Partnership for an Advanced Computing Environment (PACE), a high-performance computing resource for the Institute. Researchers also report using the computing resources of the Department of Defense or NASA.

Respondents do not necessarily conduct their research exclusively as either an experimentalist or a theoretician. While there are some that do their work employing only one method or the other, there are many who incorporate elements of both in their research. Regardless of the methodology, the management and storage of research data produced by modeling or experiment presents unique opportunities and challenges.

Much like other STEM fields, AE is increasingly a collaborative and interdisciplinary endeavor. AE faculty collaborate with other AE researchers within Georgia Tech School of Aerospace Engineering, as well as with AE faculty at other research universities. Collaborations also occur with other disciplines and schools within Georgia Tech, as well as with other disciplines and schools at other research universities. Collaborative research is the norm at Georgia Tech in the field of AE. As one faculty member respondent put it, "[T]his is the trend. There's a big push to collaborate across communities."

Scholars from AE collaborate with researchers from other science and engineering schools within Georgia Tech. Within the College of Engineering, they collaborate with faculty from schools such as the Schools of Civil and Environmental Engineering, Industrial Systems and Engineering, Electrical and Computer Engineering, and Mechanical Engineering. AE faculty also collaborate with academic researchers from the College of Computing, and with faculty from the College of Science (which includes the School of Physics, and the School of Chemistry and Biochemistry).

Georgia Tech also hosts interdisciplinary institutes where faculty from multiple colleges, departments, and individual labs collaborate around a core research area. Respondents to our survey engaged in collaborations with Georgia Tech for Robotics and Intelligent Machines (IRIM), as well as the Strategic Energy Institute. Respondents also collaborate with Georgia Tech Research Institute (GTRI).

Outside of Georgia Tech campus, respondents reported collaborations with faculty at other universities, as well as collaborations and partnerships with corporate and industrial entities. Agencies within the US government were also important collaborators, particularly NASA and the Department of Defense.

Research Data Practices and Challenges

AE faculty produce a variety of data types. These can range from simple numerical data from a sensor reading, to plots of a flow or trajectory, to extremely high quantities of high-resolution image data, to 3-dimensional data points in a simulation. An experiment or a modeling run can generate anywhere from a few megabytes or gigabytes of data (for simpler numerical data and code) to terabytes of data (in the case of image data or 3D data from simulations).

For some faculty, the product that is produced from research is a design, obtained from existing government or corporate data, rather than new raw data itself.

When it comes to saving research data, there were respondents who reported that their primary strategy was keeping it and managing it locally (and had been for some time). Local storage options used by AE faculty included personal PC's and laptops, personal external hard drives, and large capacity disc drives (such as a 20 TB drive, in one instance). Some respondents reported taking advantage of cloud storage like AWS or Dropbox, or the storage available at the campus computing center resource, PACE. Some respondents create codes that are stored in open online repositories like GitHub.

> *"We just buy external hard drive[s], and maybe after a period of time we just get rid of them."*
>
> *"[E]verything is stored in my hard drive."*

“[M]ost of my data, most of my files and everything, is in the Dropbox.”

“[W]e have a 20 terabyte disc drive, which is really not a big deal to manage.”

“GitHub is what people use[…] especially for code.”

However, there are faculty, particularly those who deal with computational simulations, who produce so much data that keeping it locally is not considered a realistic option. Data from simulations can be in such large quantities that the faculty member instead finds that analysis can only be effectively done using a snapshot of the total. One faculty member described how they could routinely generate 50 terabytes of data from a single simulation running on over 1000 processors in a computer cluster.

“[W]e run on parallel computers. So each job might run for 1,000 processors, or 2,000 processors. And each snapshot of the data, like some of the posters you can see. Either all the codes can generate – will be like 50 gigabytes, one snapshot.[…] So a thousand of them will be 50 terabytes.[…] We process the data as the simulation is going on, and write out only what we want. So a slice, a 3D snapshot, a small piece. Like a little window of the thing.”

Other faculty members that have similarly large data sets keep them on resources maintained outside of Georgia Tech, typically a government server, usually owned by NASA or the Department of Defense. But even when using government servers of massive capacity, not all research data can be kept. Said one faculty member who generates extremely large data sets:

“In Georgia Tech I have a system that [contains] 150 terabytes of drives. But there are like hundreds of drives, and they are all spinning around - but I also use NASA and Department of Defense computers that are off site. […] These are 100,000 processors, and half a million processors. Huge machines. So they have their own data storage system, robotic system. So they provide data storage. Typically when a student graduates, or something happens, we delete all the prime data, because we can't store it. Because not only it takes a lot of data to store but to access it also takes a long time. Because we're talking about a huge amount of data. It doesn't sit on one computer. It's on 500 computers. And 500 drives.”

One respondent described how data from simulations can be so overwhelming that even if keeping all the data is desirable, it’s far too big to store in the local Georgia Tech PACE computing center:

“So when we run the high fidelity simulations, we – so what we do is, we have the vehicle and then we have an area surrounding it. We call it control volume. […] So it's basically, typically around six to eight variables that we solve at every node, every point in that control volume. And then we save it. And we analyze it. And the control volume, the number of nodes, it can be, if it's a small problem, maybe three million degrees of freedom. For a larger, like a real helicopter, it can easily be 500 million.”

The respondents who had concerns about the lack of storage space for data tended to be the researchers who generated and used massive datasets (terabytes or more). Respondents that did not have to deal with such massive sets of data reported varying degrees of concern about successfully storing and maintaining data for the long term. For some researchers, the data produced by their students is only valuable in respect to their final thesis or dissertation. The publication of thesis is considered the most important repository: "they are the archival information for your data."

For some researchers, the amount of storage space was not the issue of concern, but rather it was the organization of the datasets that posed a logistical challenge. One faculty member mentioned that neither they nor their associated graduate students or collaborative researchers have the skills, resources, and time available to organize data in a meaningful way. For some AE faculty, such an organization project would not be worth the effort anyway:

> *"[If] we think the data's not really going to be used, and we're the only ones who's going to use it, that's a lot of effort to go to make all that information and organize it and then it's a waste of time if no one does it. So in the short term aspect to us, it doesn't help us. Too labor intensive on our end."*

Another researcher did not report any history of problems finding old or historical data from his lab's work, but did admit that organization could be improved, perhaps using some sort of organizational software tool or resource:

> *"[I]f you could imagine a software, for example, when I put the data there, I can add a label explaining what are some keywords. [That would] help to search it."*

There were faculty who saw the same challenges in organizing their research data in a meaningful way, but still felt that those challenges were worth meeting, and that they would benefit from having data organized and available. There were a variety of reasons stated for this viewpoint. The need to have copies of backup data in case of a technical malfunction with a storage method was mentioned, but also the need to organize to preserve access for future users.

> *"Making sure [data] are readily accessible, that's something people don't always talk about. Like yeah, great, you stored all this data in this exotic location. [Then] student number two comes along; how long does it take them to figure out where the data is located? Because it's not like a library, you go in and see it, right? What is the clear data structure by which you will store a project information?"*

The challenges of effectively storing and managing data in an organized way go beyond the lack of data management skill to do so, or the lack of time and resources to do so effectively. Organization of research data is further hindered by a lack of standards or best practices. Even if time and resources could be allocated to better organize and manage stored research data, faculty feel there are few established methodologies to follow. As one researcher put it:

> *"I see the advantages of having some standards on how the data you take is connected with the information that went along with that data. When a student writes a program, a whole bunch of things go into that data stream that says, okay, here's the data but here*

are all the settings of everything and here's the time of day. So a lot of that descriptive information, which 20 years ago would have gone in file names, you would have written a file name and said, okay, you know. So I think that a process just to do that [would be beneficial]"

Researchers in AE often must pay considerable attention to information security protocols. Respondents reported working with industry and the government in the production and use of their research data. In the case of industrial or corporate partners, the research data may be proprietary, and it is incumbent upon the researcher to safeguard the information until the corporate partner approves release. In the case of the federal government, data from a project may be restricted by law to US Citizens. Regardless of what data storage system is used, security is a constant concern. As one researcher put it:

"[Security] shouldn't be hard, but it can be difficult to make sure you know what data that a student has in their hands. Of course they sign all the waivers and all that. But something that we're working on here in AE is a student exit policy with respect to data"

At least one researcher reported having to pay out of their own research budget for a backup data security system. And even for research projects that don't have proprietary data or government-restricted data, storage cost is an issue. Faculty reported buying drives personally, having to allocate some research budget to storage hardware, and having to allocate funds to off-campus sites because local data storing options were unavailable or inadequate.

Tools for Data Production, Analysis, and Sharing

AE scholars make use of a variety of software tools for the production, analysis, and sharing of their data. The researchers use a mix of commercial software, software tools developed in-house, and open source software. Tools such as ANSYS are used to help with large and complex simulations. For those researchers whose work is more experimental, sensor data is acquired using tools like LaVision.

To analyze the data, researchers use specialized commercial analytical tools like MATLAB, Tecplot and LabVIEW, which are designed mainly for use by engineers and scientists. However, some respondents reported using even more common commercial tools like Excel for data analysis. In AE, the analysis of data often goes hand-in-hand with the visualization of the data. Tecplot and LabVIEW both help to visualize the data produced in simulations and experiments. Some respondents also indicated use of open source software, for example ParaView for data visualization analyses, and R for statistical analysis. Common computer languages used are Python, C++, MATLAB, and even the older code FORTRAN. Although antiquated, sometimes existing useful codes written in FORTRAN must be modified, necessitating some level of skill in using it.

Despite the prevalence of the use of these codes in research, there are faculty who see a need for their students to become more adept. One faculty member noted that probably "less than 5%" of his students knew Python or R, and that being more familiar with the software for data visualization and analysis would be "incredibly helpful".

AE scholars also use a wide variety of tools and platforms for sharing research data. Many faculty think of the published thesis or journal article as the public sharing of data. However, internally, data is shared through local tools like emails and shared laboratory disc drives. Data that is not sensitive is shared through document sharing platforms like Microsoft OneDrive, Dropbox, and Google Drive. Services like QNAP's Network Attached Storage (NAS) are also used for backup, storage, and transferring large data. Although sharing data internally was generally not considered a challenge by most respondents, getting large amounts of simulation data from one place to another was a problem. At least one faculty member reported that getting data from PACE to his own lab was not easy:

> *"The biggest challenge we face is actually moving data around right now. So if I want to move data, even internally, like if I want to move data, say, from PACE to, say, our lab downstairs, you know, the data transfer rates are still, for the size of the data we'd like to move, are still fairly slow."*

For faculty who use NASA or Department of Defense systems, the problem is similar.

> *"Because in the [Department of Defense] systems, once they archive it, they move it into a robotic system, another system. So if you want to pull some file out of it, you have to call it back, you know, and so that takes a while."*

In summary, research data produced by AE scholars is stored, managed, and accessed by a wide variety of means. There are varying levels of concern about the ability to keep research data for the long term. In fact, even for those faculty members whose strategy tended towards favoring a more ad-hoc approach (keeping data on laptops and external drives), funded as needed by laboratory budgets, the concern over losing data was often minimal. There were exceptions. Far more, however, expressed desires that their data be better organized within whatever system they were using. They also wished for better data transfer technology, and for their students to be better familiar with the tools available for analysis and visualization.

Future Challenges and Opportunities

Several AE scholars mentioned challenges in the AE field specifically. One scholar commented that it is getting more expensive to conduct combustion or high temperature gas related experiments on campuses since they often involve noise, smell, health and safety issues. Another challenge area is education. Scholars recognize that there are education gaps in trending technology areas, such as artificial intelligence and robotics, among engineering students. Some of the challenges are universal, especially those related to data breach or data security, and how to balance security, risk and cost.

When talking about future computing, one scholar commented that it is a challenge to get infrastructure to move data back and forth faster. Machines are getting bigger and faster.

> *"So the challenges are like as these big machines come on board, their CPUs, processing are improving, but storage is not."*

> *“You can now get 8 terabyte drives or 36 terabyte drives in one little thing. But… accessing is going to be very difficult.”*

Regarding future education and trending technology, one scholar put it this way:

> *“How do we incorporate machine learning or artificial intelligence with all of the fields? Because really big data means now you have big training sets. I don't know. In our current model a lot of those things are like housed in the robotics and the machine learning PhD program on campus. But not everybody needs a PhD level understanding of machine learning and AI. They want like an undergrad[uate] level of understanding so that they're functional, they know the terminology and they know what they're buying so that they can apply it as appropriate to their field.”*

Many scholars we interviewed consider the limited amount of time that can be devoted to research as a constant challenge.

> *This is “an era where there’s a lot happening. So really our biggest challenge is having enough time to do everything we want to do.”*

One eye-opening point raised by a couple of scholars is related to diversity. Making sure that research is inclusive of all races and genders can be both a challenge and an opportunity. It may not always be clear what inclusion means in the research process. As one asked:

> *“As we have more females and URMs (under-represented minorities) entering the areas, do we need to adapt what we're doing? What does the face of data look like?”*

Almost all the AE scholars we interviewed predicted a bright future for research in the AE field. Some exciting areas of research include space exploration, habitats on other planets, renewable energy, urban air mobility (“flying cars”), autonomous vehicles, and hypersonic airplanes and so on. As one scholar put it:

> *“It's a super exciting time in aerospace engineering right now. There's another new push for space exploration, so we have a lot of research activity in different aspects of space, whether it be creating and testing small satellites or just working out habitats on other planetary bodies, as we start to live and work on the moon and Mars.”*

> *“…in my 30 years in the industry, it’s probably the most dynamic time I’ve seen.”*

One scholar stated that multi-disciplinary teams have been formed on campus to tackle tough problems for the future of transportation, as he described:

> *“It feels like the future. It's the Jetsons, flying car kind of idea. But they're not cars, they're airplanes… just taking aviation to cities to save time and to provide greater access to transportation, and goods on demand.”*

As far as funding is concerned, AE scholars were optimistic that their research will continue to get more government and industry support.

One scholar praised what the library has done and suggested continuing efforts from the library:

> *"It will be incredibly helpful for the library to teach students data analysis and data visualization tools."*

Conclusions and Recommendations

In the course of analyzing the interview responses, the authors noted some ways in which responses by AE faculty compared to those of CEE faculty from our previous study. Several individuals in both sets of respondents mentioned that data storage has been and will continue to be a challenge in the years ahead. Another commonality was the concern with having enough time to do the research and all the associated tasks entailed by research projects. There were also notable differences. With AE respondents, we found that more individuals were involved with experiments that produce a much higher quantity of data than was typical with CEE researchers.

The analysis and findings within these responses from AE faculty led the authors to make the following recommendations for academic librarians that serve a population of AE research faculty, as well as for librarians who work in specific research data service areas.

1. *The academic research library is well-positioned to serve as a partner for data organization consultation.* While specific libraries or campuses as a whole may not be able to provide storage for research data, librarians are adept at organizing information, are familiar with standards of preservation and organization, and can help meet the challenges researchers face in the storage and curation of data wherever it may reside. Many larger academic libraries now have a dedicated research data librarian (or librarians) with expertise to provide organization consultation within the context of respective researchers' file storage and sharing limitations. To continue to meet this need, librarians need to keep abreast of changes in practice, software used, hardware used, and tendencies and habits of researchers working with research data.
2. *The academic research library is well-positioned to serve as a partner for data analysis and visualization.* In research labs, the valuable time of graduate students is often taken up learning the basics of visualizing the data in their experiments. Librarians with data visualization skills are an invaluable asset to meet this need. If a library serves a large community of engineering researchers, it would be beneficial to have a librarian with this expertise on staff. For libraries where there is already data visualization expertise, marketing to potential users is a challenging but necessary effort.
3. *Research data storage and curation will be most effective when supported at the institute level.* The need for data storage and curation will likely grow in coming years. Libraries recognize the need and the challenge of doing this effectively, but even within one discipline there are wide disparities in data size, data uses, data formats, even data access requirements. A library may try to accommodate some researcher needs with an in-house repository, but the scope of what could be collected, and the resources and funding required to maintain it, may be far greater than what many research libraries could handle. Therefore, the library should be included in discussions of campus-wide research

data initiatives along with multiple stakeholders from institute IT departments, academic departments and colleges, research laboratories, and campus research administration.

Additionally, libraries should be aware of diversity needs at all steps of the research process. One finding from our interview analysis concerned diversity, and how more under-represented minorities can enter the research arena. One faculty member in particular noted that the perception of who does “data science” may be biased in ways that unintentionally exclude under-represented minorities. Libraries should acknowledge that patterns that have been established in engineering research need to pro-actively include participation by individuals from under-represented groups. More research could be done to ensure that library services and resources to researchers do not have an inherent, if unintentional, exclusion of potential participants.

In closing, it is worth noting that this was a limited study on a population at a single research institution. Conclusions drawn and recommendations made here may not necessarily apply broadly in similar contexts. In the future, similar studies to this can be done with other disciplines. Research data services within various engineering disciplines can vary, so librarians that are engaged and involved with their research community can better meet the needs of the various engineering sub-disciplines. Rapid changes in technology can also necessitate further research, in order to better analyze the ever-evolving practices of how researchers work with data.

References

[1] D. Cooper, et al., "Supporting the changing research practices of civil and environmental engineering scholars." Ithaka S+R, New York, 2019. Available: https://doi.org/10.18665/sr.310885

[2] S. M. Samuel et al., “Analyzing data management plans: Where librarians can make a difference,” in Proc. 122nd ASEE Annual Conference and Exposition: Making Value for Society, Seattle, WA, 2015.

[3] H.L. Coates, J. Carlson, R. Clement, M. Henderson, L.R. Johnston, Y. Shorish, “How are we measuring up? Evaluating research data services in academic libraries,” Journal of Librarianship & Scholarly Communication. 6, pp. 1-33, 2018. Available: doi:10.7710/2162-3309.2226.

[4] A. Goben, T. Griffin. “In Aggregate: Trends, needs, and opportunities from research data management surveys,” College & Research Libraries. 80(7), pp.903-924, 2019. Available: https://doi.org/10.5860/crl.80.7.903. Accessed January 22, 2020.

[5] L. Perrier, L. Barnes, “Developing research data management services and support for researchers: A mixed methods study,” Partnership: The Canadian Journal of Library & Information Practice & Research. 13(1), pp. 1-23, 2018. Available: doi:10.21083/partnership.v13i1.4115.

[6] A.L. Whitmire et al., “Variability in academic research data management practices,” Program: Electronic Library and Information Systems, 49(4), pp. 382-407, 2015.

[7] E. Barsky (2017, July). “Three UBC Research Data Management (RDM) surveys: Science and engineering, humanities and social sciences, and health sciences: summary report.” Available: doi:http://dx.doi.org/10.14288/1.0348719

[8] K.G. Akers, J. Doty, “Differences among faculty ranks in views on research data management. IASSIST Quarterly, 36(2): pp. 16-20 2012.

[9] J. Carlson. (2011) “Demystifying the data interview: Developing a foundation for reference librarians to talk with researchers about their data,” Libraries Research Publications, Paper 153. Purdue University. Available: https://docs.lib.purdue.edu/lib_research/153/

[10] S.Van Tuyl, G. Michalek, “Assessing research data management practices of faculty at Carnegie Mellon University,” Journal of Librarianship & Scholarly Communication. 3(3), pp. 1-30, 2015. Available: doi:10.7710/2162-3309.1258.

[11] S.W. Parham, J. Bodnar, S. Fuchs, “Supporting tomorrow’s research: Assessing faculty data curation needs at Georgia Tech,” College & Research Libraries News, [S.l.], 3(1), pp. 10-13, 2012. Available: doi:https://doi.org/10.5860/crln.73.1.8686.

[12] C. Wiley, W.H. Mischo, “Data management practices and perspectives of atmospheric scientists and engineering faculty,” Issues in Science & Technology Librarianship, (85)1, 2016. Available: doi:10.5062/F43X84NJ.

[13] C. Wiley, “Data sharing and engineering faculty: An analysis of selected publications,” Science & Technology Libraries, 37(4), pp. 409-419, 2018. Available: doi:10.1080/0194262X.2018.1516596.

[14] J.H. Holles, L. Schmidt, “Graduate research data management course content: Teaching the Data Management Plan (DMP),” in ASEE Annual Conference and Exposition, Conference Proceedings, v 2018, 125th ASEE Annual Conference and Exposition, June 2018, Salt Lake City, UT. American Society for Engineering Education, 2018.

[15] W.H. Mischo, C.A. Wiley, M.C. Schlembach, H.J. Imker, “An integrated data management plan instructional program,” Paper presented at 2017 ASEE Annual Conference & Exposition, Columbus, Ohio, June 2017. Available: https://peer.asee.org/27572

[16] A.M. Cox, M.A. Kennan, L. Lyon, S. Pinfield, L. Sbaffi, “Maturing research data services and the transformation of academic libraries,” Journal of Documentation, 75(6) pp.1432-1462, 2019. Available: doi:10.1108/JD-12-2018-0211.

[17] J. Corbin, A. Strauss. “Basics of qualitative research: Techniques and procedures for developing grounded theory. Third ed., 2008. Los Angeles, CA: Sage Publications, Inc.

Appendix 1: Email Recruitment Text

Subject: Invitation for Georgia Tech Library’s study on Aerospace Engineering Scholars’ Research Data

Dear [First Name]:

Georgia Tech Library is conducting a study to examine the research support needs of scholars in Aerospace Engineering here at Georgia Tech. Would you be willing to participate in a one-hour (maximum) interview to share your unique experiences and perspectives? Your perspective as a scholar [insert personalizing information] will be incredibly helpful toward identifying and developing research data supports for scholars. The results will be used towards informing future efforts to improve Georgia Tech library services. Thank you so much for your consideration.

Sincerely,

Librarian’s Name

Appendix 2: Informed consent

Written Informed Consent Form

This consent form asks you to take part in a research study. The study is being conducted by [Librarian's name] (Email address 1) and [Librarian's name] (Email address 2) of the University Library.

Title of the research study: Research Support Services Study for the Field of Aerospace Engineering

Reasons for the study: This research study seeks to examine the research practices of academics in Aerospace Engineering in order to understand the resources and services these faculty members need to be successful in their teaching and research.

What you will be asked to do: Your participation in the study involves a 30-minute audio-recorded interview about your research practices and research data support needs as an Aerospace Engineering scholar. Your participation is completely voluntary. You are free to withdraw consent and discontinue participation in the interview at any time for any reason.

Benefits and Risks: There are no known risks associated with participating in this study. Subjects may experience benefits in the form of increased insight and awareness into their own research practices and needs.

How your confidentiality will be maintained: If you choose to participate, your name will not be linked to your interview responses or workspace photographs at any time. We do not include your name on any of the interview data and there is no link between this consent form and your responses.

Questions? You may contact the researchers at any time if you have additional questions about the study, or, if you have any questions about your rights as an interviewee, you may contact [Georgia Tech IRB office].

Your participation in the interview provides your consent.

Consent Form Approved by Georgia Tech IRB: June 28, 2019 - Indefinite

Appendix 3: Semi-structured Interview Guide

Research Data Support Services for the Field of Aerospace Engineering

Research focus and methods

- Describe your current research focus and projects.
- How is your research situated within the field of Aerospace Engineering?
- Does your work engage with any other fields or disciplines?
- What research methods do you typically use to conduct your research?
- How do your methods relate to work done by others in Aerospace Engineering [and, if, relevant in the other fields you engage with]?

Working with Data

- Does your research typically produce data? If so,
- What kinds of data does your research typically produce? [prompt: describe the processes in which the data is produced over the course of the research]
- How do you analyze the data? [e.g. using a pre-existing software package, designing own software, create models]
- How do you manage and store data for your current use?
- Do you use any other tools to record your research data? [E.g. electronic lab notebooks]. If so, describe.
- What are your plans for managing the data and associated information beyond your current use? [e.g. protocols for sharing, destruction schedule, plans for depositing in a closed or open repository]
- Have you encountered any challenges in the process of working with the data your research produces? If so, describe.
- Are there any resources, services or other supports that would help you more effectively work with the data your research produces?
- Does your research involve working with data produced by others? If so:
    - What kinds of data produced by others do you typically work with?
    - How do you find that data?
    - How do you incorporate the data into your final research outputs? [e.g. included in the appendices, visually expressed as a table or figure]
- How do you manage and store data for your current use?
- How do you manage and store this data for your current use?
- What are your plans for managing the data beyond your current use?
- Have you encountered any challenges working with this kind of information?
- Are there any resources, services or other supports that would help you more effectively work with the data produced by others?

State of the Field and Wrapping Up

- What future challenges and opportunities do you see for the broader field?
- Is there anything else about your experiences or needs as a scholar that you think it is important for me to know that was not covered in the previous questions?