

# **Understanding the research practices and services needs of civil and environmental engineering researchers - a grounded theory approach**

Xiaoju Chen, Jessica G. Benner, Sarah Young, Matthew R. Marsteller  
University Libraries, Carnegie Mellon University, Pittsburgh PA.

## **1. Introduction**

### *Motivation*

Researchers have different needs and behaviors during the life cycle of their research. Understanding the behaviors related to a suite of research-related activities can help librarians develop deeper knowledge about the subject fields in which their users are involved. Library spaces and services must be designed in ways that take the needs and behaviors of these researchers into account. Research behaviors and needs vary across disciplines. To address specific needs from a certain discipline, it is important for subject librarians as well as information specialists to understand the research life cycle within that discipline. In this study, we aim to understand the research practices and needs of Civil and Environmental Engineering scholars. We specifically focus on the faculty members who are conducting research in Civil and Environmental Engineering at Carnegie Mellon University.

### *Civil and Environmental Engineering at CMU*

At Carnegie Mellon University, most Civil and Environmental Engineering researchers are affiliated with the Department of Civil and Environmental Engineering (CEE), which is part of the College of Engineering at CMU. The department offers a number of programs and areas of study, including Bachelors of Science (BS), Masters of Science (MS), and Ph.Ds. It has established research strengths in a number of research fields, including sustainable, intelligent infrastructure systems, water and air quality science and engineering, resilient engineering materials, and environmentally sustainable engineering practices. These research fields are organized into three major focus areas in CEE: Advanced Infrastructure Systems; Environmental Engineering,

Sustainability; and Science, Mechanics, Materials, and Computing. The CEE department is ranked 8th by the 2018 US News and World Report for both graduate and undergraduate Environmental Engineering programs; it is also ranked 10th and 12th for graduate and undergraduate Civil Engineering Graduate Programs, respectively.

Apart from the CEE department, some employees in the Department of Engineering and Public Policy (EPP) at CMU also conduct research in Environmental Engineering. These researchers often hold dual-appointments in the CEE and EPP departments. The EPP department offers B.S., M.S. and Ph.D. degrees in various disciplines. Environmental Engineering research in the department focuses on many aspects including energy systems, climate and environment, and risk analysis.

CMU is one of 11 universities conducting similar explorations of CEE research practices. The overall study leadership and compilation is being coordinated by Ithaka S+R. Ithaka S+R is a not-for-profit organization that provides research and service guidance to help academic and cultural communities better serve the public. This project studying CEE is the latest in a larger series of studies by Ithaka S+R on the research support needs of scholars by discipline. The remainder of the report discusses our findings related to civil and environmental engineering researchers at CMU. For more information about the larger study by Ithaka S+R or to access previous studies of research support needs, you can visit the following webpage: <http://sr.ithaka.org/our-work/research-support/>

## **2. Methodology**

This study used a grounded theory guided approach consisting of the initial and focused coding steps. We recruited representative researchers from the CEE and EPP Departments aiming to balance for gender and years of experience. We sent recruitment emails in three waves and ultimately interviewed nine out of 27 possible researchers. Recruited researchers participated in a 60-minute interview conducted in their office. The semi-structured interview protocol (see [Appendix A](#)) was designed by Ithaka S+R. Once the interviews were transcribed, we conducted two rounds of analysis. First, each team member coded all interviews and generated open codes.

Next, we discussed our codes and collated them into a set of four themes. Each team member then coded each interview for one of the four themes. For more details about our methodology, see [Appendix B](#).

### **3. Findings**

#### 3.1 “Collaborations, connections, interactions”: CEE research depends on a diverse research network

Collaboration was a theme common to all of our interviewees and clearly plays an important role in civil and environmental engineering research. None of the researchers we interviewed work alone. We found that a rich landscape of collaborations, partnerships, networks and consortia were critical to research at nearly every stage in the research life cycle. The motivation to establish research collaborations, and the diversity of sectors involved, varied. In most cases, collaborations were driven by either a need for expertise or a need for access to data. The following two sections discuss these different types of collaborations.

##### *Collaborating for expertise and across disciplines*

Research collaborations with other faculty on campus, usually in different departments, were most often established to add expertise or technical skills to a project. As stated by one researcher: “[E]verybody has their specific expertise, and so you bring that to the table...It’s really kind of fun to solve a problem that could not be solved by a single discipline”. Another researcher discussed their efforts to recruit faculty collaborators from machine learning to develop tools to improve the efficiency of preparing large raw datasets for analysis. At the same time, some collaborations were borne of the need to share expensive equipment or through connections via large research consortia. Regardless, the multi-disciplinary nature of CEE research was evident and reflected in the many faculty collaborations spanning multiple departments, as well as the more formal groups such as research institutes, of which there are many at CMU. Collaborations with colleagues in public policy, economics, business, design, robotics, chemistry, machine learning, decision sciences, and statistics were all mentioned. In addition, dual or more

appointments in different departments were common amongst our interviewees. Notably, many of our interviewees viewed themselves not as engineers, but as specialists in other disciplines, such as economics, public policy or computer science. This reflects CMU's strong emphasis on fostering interdisciplinary research and the need for an interdisciplinary perspective to solve real-world problems. Some researchers highlighted challenges associated with interdisciplinary research. For example, two noted the disciplinary language barriers that may exist in the early stages of collaborating with someone in a different discipline. However, communication, face-to-face interactions and collegiality were all mentioned as important factors to a successful collaboration.

### *Establishing partnerships for access to real-world data*

Very close in importance to academic research collaborations were partnerships with industry and government organizations. The research in CEE at CMU is applied research, focused on solving problems in the real world and helping practitioners and policy-makers make better decisions. Thus, working with industry and government was common amongst our interviewees. In some cases, researchers worked closely with these partners to identify pain points and needs for tools to improve efficiency or make more informed decisions. More often, these partnerships were key to accessing interesting real-world datasets. Today, industry and government agencies collect vast amounts of data, but may not have the capacity or expertise to analyze it. Thus, these partnerships are mutually beneficial to both parties, with industry and government benefiting from the models and tools generated by the researchers. One researcher put it this way:

[Our industry partners] understand the capabilities of technologies without needing to commit a lot, and I get an opportunity to be at the real world and understand the issues of the real world.... So, it's not like their goal is to come up with a new algorithm to do something, right? Which would be my goal. But their goal is to kind of learn something about the technology without truly investing in a technology.

### *Finding collaborators*

The ways that people seek collaborators and partners for research varies. Interestingly, several researchers mentioned simply using Google to find researchers with a particular area of expertise or with interesting data. On the other hand, researchers also rely on their existing networks to identify new collaborators. Another researcher discussed attending conferences for state and local governments as a way of establishing connections for potential future research and learning about real-world challenges that aren't necessarily reported elsewhere:

[Y]ou need to go [to these conferences] to meet those people, for them to say hey, you did a great study for Pennsylvania. Will you do one for us, and we'll give you the data. [That's] where you'll see the conference presentation that says 'Well, here is what the problem is in our state'. There is never a conference paper. There is never a journal paper.

Once established, a successful collaboration or partnership can last many years, yielding numerous research projects and publications.

### *Challenges of collaborating*

While most acknowledged that collaboration is essential, challenges around collaborating and partnering with industry or government were also discussed. Since most of this work revolved around access to existing data (or in some cases collaboratively generating data), data management and sharing were mentioned as collaboration challenges, though not frequently and only in general terms. More often, researchers mentioned challenges with the writing and publishing phase of collaborative research. One researcher discussed the challenges of collaborative writing and reference management, an area that the library currently provides support in with tools like Mendeley and Zotero. Negotiating authorship order, in a way that favored their own graduate students while still recognizing the contributions (often the data) of co-authors at other institutions or organizations, was mentioned by one interviewee. Another interviewee who worked closely with government organizations discussed the fine balance of maintaining good relationships while publishing something that might be viewed as critical of a partner organization.

### 3.2 Data practice and research cycles

#### *Civil and Environmental Engineering research cycles*

Civil and Environmental Engineering researchers conduct research using data and analytical models they develop to solve real-world problems. There are two typical research practices in terms of collecting data. One practice is to collect data produced by the researchers themselves, either by experimentation or building data models. The researchers develop analytical models to identify what data should be collected, and then design experiments to collect these data. The other practice is to use data generated by others. The researchers locate and collect data produced by third parties and manipulate these data for their research. These two different practices heavily affect the researchers' data-related behaviors, from data production or access to data sharing, data processing, and management.

#### *Producing and accessing data*

The researchers who use third party data need to locate the data before the analysis starts. Therefore, they spend more time locating data and face more challenges in accessing data. Some of these researchers consider the process of locating data similar to the process of doing literature reviews - they simply use web searches (often just a Google search) to locate data. To access data that are not published, the researchers tend to directly approach potential data holders. According to some researchers, the field “is not big” and they are “aware of data because of interactions with other researchers.” However, some researchers mentioned that one challenge for locating data is that “there’s no one repository where all the data is,” it’s a challenge to find data “unless you know where to find it.”

Data produced by researchers are often free. In some cases, researchers pay for datasets produced by government agencies or private sectors. Occasionally some databases are too expensive and researchers have to find alternative solutions. Some databases are not even available to academic users because the database developers fear that “you’ll just give it away.” Sometimes this is the

biggest challenge for accessing data. According to one researcher: “You just have to realize there’s some times where if somebody just got data and it’s just what you need, you’re just not gonna get it.”

Despite the challenges, accessing data leads to future collaborations. One researcher mentioned that he would ask for a small sample of data from the producer, and show the capability of his research. After building the initial collaboration, the researcher asks for more data and starts a research project with the data provider.

### *Data sharing*

Some researchers believe that in the Civil and Environmental Engineering field, sharing raw data is not as progressive as in some other fields, such as Robotics. Most researchers do not tend to share data unless required by journals or funding agencies. Occasionally they share raw data with other researchers via one-on-one requests and believe this type of sharing is “based on good faith.” There are some discussions and agreements on how the raw data are going to be used. For example, one researcher said: “If people ask me..., I have a conversation and then...we can agree to use....[W]e have kind of an understanding of how that’s gonna be used.” The researchers also communicate with the requestors about the limitations of the data such as the assumptions and conditions when the data were collected.

This data sharing practice shows one of the many reasons why the researchers are concerned about making their raw data publicly available: they do not want anyone to misinterpret their data and draw wrong conclusions. For example, one researcher said: “Some people won’t like that idea [be]cause opening it up to anybody in the public ...[allows anybody] to poke around through your stuff.” Another researcher mentioned:

I can't guarantee the accuracies of the data ..., but when I'm kind of one-on-one sharing, I can say, "hey. Here are the conditions that we developed. This was the scanner we used. I don't even know how accurate it's gonna be for whatever worth, take it." But, when you

put it out there, I feel like I need to have all that either guarantees or some specifications of what the data accuracies and other things mean, and that's kind of scary to me.

There are some other reasons for not sharing raw data. Some researchers believe that the data they produce is part of their research power:

I have done [data sharing], ... I don't even know why. Because, I see how great it is when ... for example, the CS (Computer Science) folks do it – but for some reason I don't see it in our community, that everybody feels like whatever data set they own, it's kind of power for their research.

Another main reason is that sometimes their data involves human subjects and cannot be shared. Researchers in CEE occasionally conduct research that involves human subjects, for this type of studies, researchers understand the policies and do not share data unless it is approved by the Institutional Review Board.

When researchers do decide to open their data to the public, they use different methods and tools. The researchers share their data by putting their data in some online data repository, such as GitHub, Zenodo or some other online platforms, or host a server on campus. Some researchers also share their models and results online. For this type of sharing, the server's capacity, cost and maintenance efforts are the three main challenges. Researchers do not share raw data that are produced by third parties, because they do not know if the data can be shared or if there are special requirements from the data producer in terms of sharing these data.

### *Data processing*

The researchers who analyze data produced by others often do not consider themselves as “creating data”. They do not consider the model or model outputs they developed as their data. For example, one researcher said: “Mostly we are working with the data source itself and almost accidentally creating data.” The researchers process data produced by others and use interdisciplinary research methods to build analytical research models for their studies. These data are often in various

formats and are gathered from different data sources such as government agencies, companies, or other research institutes. For example, some researchers use data on vehicle makes and models, others analyze social media records, and still others use descriptive data such as Likert scales.

On the other hand, the researchers who collect their own data are often more constrained to the Civil and Environmental Engineering field and often collect data in real-world projects. For example, some researchers apply sensors and collect data to analyze the performance of infrastructure.

All researchers process their data by developing analytical models using either off-the-shelf tools such as GIS software, MS Excel, or developing their own scripts with programming software, such as Matlab, R, Python or C++. Most of the researchers agree that data cleaning is the most time-consuming process. In addition, the cost for processing big datasets is often expensive. Some researchers mentioned that although the university has resources to process big datasets, such as high-performance computing machines, there is no instruction on how to use these resources. They appreciate any educational materials such as a data processing course or a workshop for first year students. Some researchers think that converting between data formats is another challenge in processing data because it is very time-consuming and there is a lack of effective tools.

### *Data management*

The researchers deal with different sizes of data, which can vary from small (20-30 entries) to “enormous” (terabytes of data per week). The researchers who deal with big data consider themselves as non-traditional civil and environmental researchers and are more aware of data management issues. The researchers who deal with somewhat smaller datasets do not consider data management a big issue; some researchers believe they do not have any data management problem at all. Others who only use third party data associate data management more with data processing, such as how to more efficiently manipulate data.

Several researchers rely on their students to manage data for ongoing projects. Some mentioned that they had problems because of this practice. For example, one researcher said: “Students’ data

management plan is “ad hoc”, such as they don’t document their codes.” and another researcher mentioned: “We’ve actually had two graduate student laptop failures that have caused some problems.”

Some interviewees mention that they have a lab server, and they require all students to upload their work to the server, but this policy is not fully implemented. To avoid data loss due to graduation, students were also asked to submit a copy of their data before they graduate. Despite this practice, some researchers still believe that when students leave, the maintenance of their shared data becomes an issue. For example, one researcher said: “I feel like we’re kind of losing the momentum with every student leaving, and then we’re building it up again.”

Using hard drives to store data is a common practice; however, the researchers understand that in the long term it is going to be an issue. One mention that they have lost data because of hard drive failures. Some researchers mention that they have used the university’s servers, or the library’s institutional repository, to store their data; however, most of them think that the space provided is not enough. Backing-up data is a common practice among the researchers, the back-ups are often stored on hard drives or with online platforms, such as Box and Dropbox.

Most researchers, whether or not they are satisfied with their data management practices, discussed some potential services that the library or university can provide to improve these practices. They expressed interest in having a good data archive platform provided by the university. One is interested in using this platform to archive any type of data generated by their research projects, including emails, unpublished reports, and phone call records. Some mentioned that they wish this platform will also help them maintain the data in the long run: “I want to just put it over there and not worry about it”. The researchers are also interested in getting help from the university to draft a good data management plan for submitting research proposals, but they believe that there is little help currently. The cost of managing data is also a general concern among the researchers. Some claimed that they just use hard drives to store data because it is cheaper than using cloud services for large datasets.

### 3.3 How Civil and Environmental Engineers Interact with Published Information

#### *Information Access*

Numerous studies of the information seeking behaviors of engineers have been done in the past. While this study contained some echoes of these earlier works, there are important differences that need to be addressed. First, this study is of faculty engineers and the information seeking behaviors of practicing engineers could be different. Thus, any comparisons must keep this in mind. Second, with the passage of time, the variety of tools available to the civil and environmental engineering faculty have evolved. The most noted practices we encountered echoing earlier studies were asking their colleagues for information, using the cited references of known literature and locating citing references with tools like Web of Science and Google Scholar.

Some interviewees indicated that their interaction with published information has changed to more of an editorial role of the work of their graduate students. Literature reviews for research projects were primarily undertaken by graduate students with the faculty member evaluating and reviewing for quality. As one researcher put it: “Well, in my case I rely mostly on my students. ... I ... push them to kind of go beyond whatever is published in the recent journals.

“... research in 1950s, sometimes '70s, just basic ideas, foundational work, that you can build off of. And, they are surprised ...”

There was one intriguing story that involved locating grey literature. The interviewee divulged that government employees would have internal reports that contained needed information, but the government didn't want the items to be cited: “... the states don't want us citing these documents because then more people are going to start asking for them and stuff. ... We're just trying to be the first paper to, in the open literature, go on record to say what this number is.”

Is it possible that government employees felt that dissemination of reports was too difficult of a task? Or, was there just a fear of having a mistake discovered? The interviewee commented further:

[...] it's hard to do the forward and backward kind of searching through citations when nothing is in the open literature. And it's not like the results of these studies are like inflammatory. It's not like they're showing that these programs have high or low failure rates. They just don't have to publish them, and so they don't.

When asked about ever getting unpublished reports, the interviewee stated:

We kind of got people in trouble, because we looked at the report. We read the first page and realized there was a big error on the first page that they had already submitted, and it was like they said the failure rate was 125 percent ... So, they said, well hold on. Let us get back to you and I get a call like an hour later ... yeah sorry, that's a typo.

Another interviewee also mentioned difficulties with documents at state agencies. In that case, the state was still producing reports in print format only and interested parties had to read them at the agency and pay for photocopies.

### *Information Management*

There appears to be a patchwork of tools being used for managing references with EndNote, Zotero, Mendeley and BibTeX in use within the group of nine interviewees. It's important to note that the institution has Mendeley for Institutions. However, there was just one definite use by an interviewee, another interviewee's response was vague but could be a Mendeley user, one interviewee's prior students were Mendeley users, but not lately, and one has a co-author that wanted to use Mendeley. One interviewee mentioned being a huge EndNote fan that even searched databases via EndNote. One interviewee used a combination of Zotero and BibTeX. Two interviews didn't cover the topic.

### *Information Trends*

Conference attendance was mentioned numerous times. One researcher described his activity: “I go to a lot of conferences and I travel like a maniac.” Even when some researchers mentioned difficulty with affording conferences, one indicated that they would review for conferences to help them keep up with ongoing research efforts in their field.

Twitter as a tool got mixed reviews. One researcher stated “I expect Twitter to disappear in the next five years, but I may be wrong.” However, another researcher mentioned: “Twitter is powerful for keeping up with what people are doing (much more so than reading), someone may know you, but not the eight or so colleagues you’re discussing your work with.”

Three researchers mentioned that they use the Google Scholar update feature to keep track of emerging literature. Interestingly one mentioned using the feature but still felt less up-to-date. One mentioned that the feature was his primary method of staying up-to-date.

One researcher mentioned tables of contents from publishers, but only used them when he had time.

### *Information Types*

Peer-reviewed journal articles were commonly mentioned as being highly valued by the researchers interviewed, but conference papers seemed to be important only to those whose work delved into areas of computer science. Four researchers mentioned using books with one explaining that they were helpful when venturing into a new area. A couple mentioned the use of grey literature and one mentioned the importance of patent literature. One researcher mentioned the use of policy documents and government reports from entities such as the Government Accountability Office and the Nuclear Regulatory Commission.

### 3.4 Scholarly outputs: traditional patterns prevail

*Venue selection: where should I publish my paper?*

For a majority of the interviewees, the most important criteria for selecting where to publish an article is the quality of the journal. Quality in these cases is measured by impact factor or other bibliometrics. The second most referenced criterion was the audience that the researcher is trying to reach. One researcher explains the relationship between these two criteria: “[W]e work our way down the prestige ladder and decide based on how general we think that the findings are, or how specific, you know, what readership we want basically.”

Another researcher describes why they prefer to publish in a smaller set of journals: “... they all have different rankings, too. ... If it’s a really star paper, I refer it to a certain journal, and maybe it’s not so highly ranked in my mind, might be another one. So I feel more comfortable, more at home publishing in the same journals.” They further elaborate on how frequent interaction with a particular journal increases their knowledge about the audience of the journal which in turn leads them to feel more comfortable submitting to those known venues:

“Once you start publishing a certain journal, you kind of get to know the audience better and you feel much better and much more comfortable sending - you know which angle you need - what they expect, what they don’t, how it fits. You know, I can look at a paper and I can say, “Oh yeah, this is more [journal A]”.”

The impact of audience is more important for those faculty who have dual affiliations. One researcher whose work engages both civil engineering and computer science disciplines, notes that the impact of the venue or how the “venue is perceived by the community” is important but the more pressing question for them is:

“Who is the community that am I targeting? Since I work on an intersection of fields, sometimes the work that I’m doing or the result that I have is more interesting from a computer science

perspective, then I publish in the computer science conferences. If it's more interesting to the engineering then I go for an engineering journal.”

This same researcher noted that they send the draft paper out to their professional network and ask where they would publish it. This may be related to their affiliation with computer science and the volume of conference venues as opposed to journals. Other researchers mentioned a fourth (in some ways data-driven) way to identify a venue by looking at the paper once it is complete and listing which journals or conferences appear most often in the works cited. This seems like a very clever way to identify the fit for a particular paper or set of results.

In terms of the types of publications, some researchers have a preference for professional society publications over for-profit publishers. Those researchers who are conducting more traditional civil and environmental engineering research preferred peer reviewed journals exclusively, publishing only abstracts in conferences for networking and feedback purposes, while those conducting research overlapping with computer science, where conferences are often the final destination for research, described publishing in both conferences and journals. For example, one researcher explains “I don't work only on civil or only on engineering. In the computer science world, in some disciplines, the conferences are the final place for high impact work, not journals.”

#### *Open access publishing: breaking up is hard to do*

The researchers had a wide variety of experiences with open access publishing. One researcher had published a fully open textbook, two others had participated in a journal using open reviews, and almost all had published one open access journal article. In fact, many of the researchers expressed a desire to publish in open access venues more often. One felt that all publications should be open because “science is for humanity”. Another focused on the exclusion of “the whole developing world” if publications are not open. Many of these same researchers and a few others simply felt that open access publishing has a larger reach for their work. In other words, more people can read and use their research. In fact, this is the main criterion one researcher uses to decide whether to pay for sending an article to an open access venue, “They charge you to make

the article open access. And I've paid for that a number of times when I thought the paper was going to be something that, you know a good publication, like the one that won the award.”

While these conversations indicate that some of the researchers understand the benefits of publishing in open access, many of them continue to publish in traditional venues. Researchers expressed several reasons for this, including cost, and low or unknown quality of the open access venue. The most frequently cited reason was the cost. Several indicated that the fees are “expensive” or “costly right now” and others discussed restrictions on article processing charge (APC) funds available from the library. One researcher noted that they are only for “venues that don't double dip” while another felt frustrated because, [Y]ou have to meet certain criteria to [get the funds]. We typically don't meet them.”

At least one researcher felt that everyone they want to reach with their publication would be able to gain access via traditional channels. Still others would not pay to publish an open access article in a traditional journal opting only to publish open articles in fully open access journals. For others, these fully open journals represented middle to low quality venues or published articles of mixed quality. For example, for some areas, the top journals are not open access, “[T]he one open access journal in my domain is not a top journal ... it's kind of medium/ok journal, so I don't publish my paper to that.”

As we saw in the section above, the quality of the venue was the most important criterion for deciding where to publish. In several cases, researchers did or would pay for open options if the quality of the venue was high; unfortunately, the perceived quality of open access options in CEE are not high enough for researchers to stray from the traditional high-quality venues. A related issue expressed by at least one researcher is that junior researchers who want to gain employment must publish in top journals. Given the perception of many of the researchers (and others in the field) that the quality of open access venues is low or mixed, these are not currently considered good or respected journals. For example, one researcher discussed the impact on graduate students,

“[W]hen you have graduate students at such a high-level at a university, and they want to go out and get jobs at other high-level universities, you are gonna have to be published in the journals

that everybody respects in that field. So you can't really take a student and say well let's publish this one in the open access journal because it is open access."

In a field where the top journals are not open, this means that traditional venues prevail as a measure of quality and success in the field.

*Promoting their research: getting the word out*

Most of the researchers promote their research outputs in some way. Several researchers use press releases to disseminate their work. One researcher issues press releases when they publish in high profile journals: "... obviously when we have papers in Nature and Science, we do press releases, and we produce some videos..." This same researcher also uses another method to communicate their work to colleagues. In the case that their publication was in an uncommon venue, they send a Dear Colleague Letter to the department. Their comment indicates that this is possibly a more common practice, but we did not encounter others in this study who mentioned this promotion method.

"... every now and then there is something that wound up in some place that's a little bit obscure for one reason or the other. And those I will say to people, hey, we published this in this special issue that was a fest for someone that is retiring or whatever... there is no way your gonna read this journal, I think it is an interesting article. You might want to read it."

Writing short pieces for trade journals was also used for promotion. One researcher was asked by several trade journals about some of their research so they wrote a short piece describing the work. Many of the researchers reported maintaining websites for specific projects or for their research group as a way to promote their activities.

Interviewees were also asked about their use of social media to communicate about/promote their work. Most reported that they do not use social media professionally for a variety of reasons. One researcher felt that using social media professional was "...a little self-aggrandizing" and mentioned that "it's enough work to just get the darn things in the press." Two researchers

mentioned a generational gap and that they are from an older generation, “I think it’s typical of my generation, I think younger generation, they are definitely all over the social networks and blogs and op-eds and this and that. I’m more traditional at this point.”

There were a few interviewees who have used social media professionally; however, in our sample, these were the exception. One researcher regularly published short pieces to a blog, “I used to have a blog for six years or so... [and] I was publishing three times a week or whatever, but then I stopped. I don’t know. Life changed... I’m just publishing papers at this point.” A second researcher interviewed is very active as a producer and consumer on Twitter. This researcher uses Twitter, “[M]ostly to keep in contact with what colleagues are doing and to advertise.” Although most of the researchers are not actively using social media, many of them recognize the potential benefits of the platforms.

The researchers use new technology and services to discover and access information for their research, and process these information with methodologies they develop. The outcomes of their research are mostly shared by published in scholarly journals and some other venues. Managing information in their research cycle can be challenging, especially when researchers are dealing with “big data”; however, most researchers are aware of the services they can use and are actively seeking for better methods.

#### **4. Recommendation**

The researchers expressed their challenges and new services they can use to improve their research. In this section, we list a number of recommendations that can be used as a guide for how to improve the services from the library.

##### *Bridge the collaborations*

In the Civil and Environmental Engineering field at CMU, collaboration is significant due to its interdisciplinary nature. The researchers work with colleagues, students, partners from industry, government and academia in all parts of their research cycle, for the purpose of conducting cutting-

edge research and pushing the field forward. All of the researchers in our study indicated the important and core role of collaboration in research success and professional development. Several things were mentioned by our interviewees that present opportunities for new or expanded library services, or simply for improved outreach about existing services, related to facilitating the establishment and process of collaboration. First, the library can raise awareness about library services and expertise in tools that enable collaborative reference management for co-authored manuscripts. This challenge was mentioned by at least one researcher and while there was awareness of tools like Mendeley and Zotero, there seem to be opportunities for the library to provide training, particularly to incoming graduate students, who are often tasked with managing information for projects.

Second, the library can improve outreach and training related to tools that can enhance discovery of potential collaborators. The interesting practice of seeking new collaborations by Google-searching highlights the value of tools with built-in ontologies and visualizations designed for identifying possible collaborators. For example, the library currently provides access to Sci-Val and Dimensions, which both enable discovery of research and researchers in particular domains based on publications and grant funding. Raising awareness of such tools amongst faculty and department staff may improve the discovery of potential collaborators. Finally, the library can provide access to and training on emerging collaborative research platforms. The library recently added Open Science Framework for Institutions, and is expanding support and outreach around this tool. Tools like this, as they are more widely adopted, may ease some of the collaboration pain points mentioned by our interviewees.

### *Strengthen data practices*

Data play an important role in Civil and Environmental Engineering discipline. By collaborating with other more data-intensive disciplines, Civil and Environmental Engineering researchers have raised more awareness of the importance of data management and sharing. CMU libraries could provide necessary services to strengthen data practice in this discipline. Here we recommend the following tasks to promote more sustainable data practices.

First, the library can promote data sharing by emphasizing the benefits, such as bringing in collaboration. Current Civil and Environmental Engineering researchers have doubts and worries regarding sharing their data; promoting data sharing can be challenging. However, the library can help researchers understand the benefits of sharing their data. For example, the liaison librarians can communicate with the researchers about possible collaboration built by sharing data. The library can propose feasible data sharing plans to the researchers to make the sharing process easier and more effective. Second, the library can provide educational material, such as courses, workshops, and handbooks to introduce how to process data by using computing infrastructures on CMU campuses. This is also a good opportunity to introduce new library services to patrons. Third, the library can improve our institutional repository by extending our size and format capabilities for data deposits. During the interviews, we found that the researchers understand the benefit of using institutional repositories to store and manage data; however, our library does not provide maintenance services to the data deposited or published in the repository. To meet the researchers' need, we recommend that the library explore possibilities of providing maintenance services for deposited data. Lastly, the library should emphasize the importance of data management to faculty members and students. As mentioned by some of the interviewees, they rely on their students to manage data. It is crucial to educate the next generation of the importance of having a sustainable data management plan in their research project.

#### *Provide sufficient published information in today's technology world*

The results of the research provided valuable information on how the researchers interact with published information in today's technology world. A range of new products and platforms have been provided by private sectors and non-private organizations to researchers for better accessing information and facilitating communication. Library services should keep up with this trend and assist the researchers to take advantage of these new technologies. The library might need to have familiarity with a range of products/formats, such as EndNote, Mendeley, Zotero, BibTeX, to be able to meet the various needs in citation management. Librarians should explore the use of Twitter for monitoring research trends and consider the broader implications on scholarly communications. The use of citation databases remains important - providers of library services need to be "fluent" with Google Scholar, Scopus and Web of Science. Apart from utilizing new

technology to assist information access and management, the results of the research showed that there was a richness to the variety of literature used by our Civil and Environmental Engineering researchers, it should raise the awareness of collection building.

### *Supporting scholarly outputs*

Quality and audience were the two most important criteria to evaluate possible publishing venues for nearly all researchers in the study. Knowledge about traditional measures of quality for publishing venues is provided by the library via subscription to Web of Science & their Journal Citation Reports; however, the audience of particular venues is less known. Librarians can help researchers navigate emerging quality measures and tools such as Altmetrics to help them identify higher quality open venues and track the impact of their research beyond citation counts using metrics like the Altmetrics Attention Score. These alternative metrics can help researchers understand the best methods for promoting their work whether it is a traditional press release or a post on social media. Several researchers mentioned that they learn the audience of a journal by submitting papers to that venue. This process takes time, possibly a career, to figure out. One way the library can support increased knowledge about the audience of venues is to consult with senior CEE faculty to generate a tool/visualization that identifies the core journals CEE researchers at CMU send papers to with descriptions of their audience and expectations.

One significant point for many of the researchers was the tension between their desire to publish in open access venues and both the cost of open access publishing and the perceived quality of open access options in the field. One way our and other libraries are bridging the cost gap is to offer help with Article Processing Charges (APC). While not the perfect solution, at least one researcher mentioned that they were ineligible for the funds, these funds are commonly used by researchers on the campus to supplement the cost of open access venues. The issue of venue quality and transitioning away from the for-profit publishers is a long fight that requires disciplinary, institutional, and ultimately, societal changes. In the meantime, the library can support researcher's desires to publish openly by promoting our institutional repository as a tool to increase the reach of their research and by debunking the myth that open access equals low or mixed quality work. The library can promote the institutional repository by submitting "author's manuscript versions"

of their work to the institutional repository. Lastly, we can help debunk the myth of low quality by helping researchers navigate their open access options and identify sufficient open access options for their research area.

## **Appendix A: Interview Protocol**

This semi-structured interview guide was developed by Ithaka S+R and utilized in this study.

### **Semi-Structured Interview Guide**

#### **Research focus and methods**

- Describe your current research focus and projects.
- How is your research situated within the field of Civil and/or Environmental 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 Civil and/or Environmental Engineering [and, if, relevant in the other fields you engage with]?

#### **Working with others**

- Do you regularly work with, consult or collaborate with any others as part of your research process?
  - If so, who have you worked with and how?
    - Lab or on-campus research group
    - Other scholars or researchers [e.g. faculty at the university or other universities, student assistants, independent researchers]
    - Research support professionals: e.g. librarians, technologists
    - Other individuals or communities beyond the academy
    - Others not captured here?
- Have you encountered any challenges in the process of working with others? [focus on information-related challenges, e.g. finding information, data management, process of writing up results]
- Are there any resources, services or other supports that would help you more effectively develop and maintain these relationships?

#### **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?

### **Working with Published Information**

- What kinds of published information do you rely on to do your research? [e.g. pre-prints, peer-reviewed articles, textbooks]
  - How do you locate this information? [Prompt for where and how they search for information and whether they receive any help from others in the process]
  - How do you manage and store this information for your ongoing use?
  - What are your plans for managing this information in the long-term?

- Have you experienced any challenges working with this kind of information?
- Are there any resources, services or other supports that would help you more effectively work with this kind of information?

### **Publishing Practices**

- Where do you typically publish your scholarly research?
  - What are your key considerations in determining where to publish?
  - Have you ever made your scholarly publications available through open access? [e.g. pre-print archive; institutional repository, open access journal or journal option]. If yes, describe which venues.
    - Describe your considerations when determining whether or not to do so.
- Do you disseminate your research beyond scholarly publications? [If so, probe for where they publish and why they publish in these venues]
- Do you use social networking or other digital media platforms to communicate about your work [e.g. ResearchGate, Twitter, YouTube]?
  - If yes, describe which venues and your experiences using them.
  - If no, explain your level of familiarity and reasons for not choosing to engage with these kinds of platforms.
- How do your publishing practices relate to those typical in your discipline?
- Have you encountered any challenges in the process of publishing your work?
- Are there any resources, services or other supports that would help you in the process of publishing?

### **State of the Field and Wrapping Up**

- How do you connect with your colleagues and/or keep up with trends in your field more broadly? [e.g. conferences, social networking]
- 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?

## **Appendix B: Methodology**

## 2.1 Grounded Theory

This study used a grounded theory guided approach consisting of the initial and focused coding steps. We did not generate a theory from the analysis. Instead, we offer a set of themes that emerged from our interviews. Grounded theory is defined by Corbin and Strauss (2008, 1) as “a specific methodology developed by Glaser and Strauss (1967) for the purpose of building theory from data”. Today, grounded theory is “one of the most popular research designs in the world” (Birks and Mills 2011, 1). This is likely due to the “systematic, yet flexible guidelines for collecting and analyzing qualitative data” that grounded theory methodology provides (Charmaz 2006). According to Charmaz (2006, 46), *coding* “is the pivotal link between collecting data and developing an emergent theory to explain these data. Through coding, you *define* what is happening in the data and begin to grapple with what it means.” The process of coding begins with initial coding which is done word-by-word, line-by-line, or paragraph-by-paragraph. The next, concepts identified during initial coding are synthesized using step in the process is called *focused coding*, which results in a set of This step is about synthesizing existing concepts and identifying themes and subthemes. Themes are defined as “higher-level concepts under which analysts group lower-level concepts according to shared properties” (Corbin and Strauss 2008, 159). These themes and subthemes guide our interpretation of the data.

## 2.2 Data sampling and recruitment

In order to have representatives from researchers in all of the career stages and sufficient numbers of female researchers in the study, we started the recruiting process within a smaller sample of the whole population of 27 faculty members who were actively conducting research in the CEE and EPP departments. We sent recruitment emails to 13 representative researchers first. After receiving feedback from these potential participants, we sent a second round of recruitment emails to the remaining researchers.

From 27 faculty members, nine participants agreed to participate, including assistant professors, associate professors and full professors. Among the nine participants, the percentages of assistant,

associate, and full professors were 11%, 22%, 67%, respectively, compared to 16%, 19%, 69% in the general population. Eleven percent of the sample was female, compared to 30% in the general population.

### 2.3 Interview Process

Participants were recruited for a 60-minute interview to be conducted in their office. The semi-structured interview protocol (see Appendix A) was designed by Ithaca S+R. At the beginning of the interview, the participants were given a consent form and informed about the goals and risks of the study. Participants were also given the opportunity to consent to the interviewer taking photographs of interesting workspaces and other aspects of the participant's office. The interview consisted of questions relating to the participants' research focus and methods used, their publishing practices, how they keep up with trends in their field(s), and their experience working with other people, data, and published information. The interviews were audio recorded and later transcribed. After transcription, the audio recordings were destroyed. The transcribed interviews were used in our analysis and an anonymized version of the transcripts were shared with Ithaca S+R.

### 2.4 Coding and Analysis

An initial round of open coding was carried out by all team members on all interview transcripts. We then grouped open codes and used these as the basis of discussion to determine key themes. We arrived at a consensus on four main themes: interactions with published information, interactions with data, collaborations and partnerships, and scholarly communications. We developed a codebook including sub-themes and inclusion and exclusion criteria for each theme. Each theme was assigned to an individual team member. We then revisited all of the interview transcripts, conducting focused coding with our assigned themes in mind and with the codebook as a guide.

Once all interviews were coded for all four themes, we discussed our findings as a group. We highlighted unexpected findings and the absence of things we had expected to hear from our

interviewees. We noted points of overlap among themes as well as points that were commonly mentioned among our interviewees. Based on these discussions and our focused coding, we developed the following report of our findings.

## References

Birks, M., & Mills, J. (2011). Essentials of grounded theory. *Grounded Theory: A Practical Guide*, 1–14.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative research*. London: SAGE Publications.

Corbin, J., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3rd ed.). Los Angeles, CA: Sage Publications, Inc.