ASSOCIATION OF RESEARCH LIBRARIES

Library Impact Research Report

Discovery of Published Information by Early-Career Science Faculty

Research Leads: Fern Brody, Judith Brink, Anne Koenig, Dan Przybylak, and Berenika M. Webster

University of Pittsburgh Library System

November 14, 2022

Association of Research Libraries 21 Dupont Circle NW, Suite 800, Washington, DC 20036

(202) 296-2296 | ARL.org

Table of Contents

Introduction

This report is part of collaborative work done by the University of Pittsburgh and the University of Washington to understand the role of libraries in supporting research productivity in their institutions. The Pitt team focused on the discovery stage of the research process, with a specific focus on early-career researchers in the hard sciences.

A traditional role for an academic library that supports research and doctoral programs has been to provide access to discovery databases—book and journal content. Historically, collection size also served to help define the prestige of the university. In fact, the subscription to a specific database or access to a core set of journals was often used as an effective recruitment tool for desirable researchers. However, shrinking library budgets and escalating costs for journal subscriptions, along with dramatic changes in scholarly publishing, have put into question the importance of the library's role in providing gateways to content discovery. In FY19, the average collections budget for an ARL library was $14,578,434, and the last decade has seen a steady decrease in allocations to collections budgets.[1] On average, academic libraries spent 76% of their materials budget on ongoing subscriptions. At the University of Pittsburgh Library System (ULS), 78% of the materials budget was spent on subscriptions to journals and abstracting and indexing (A & I) databases.[2] In FY19, with annual subscription rates of science databases ranging from $20,000 to over $200,000, the value of these resources has become a concern for library administrators.

Dovetailing with the high costs of subscriptions is the shifting landscape in scholarly communication and disruptive influences from nontraditional sources. Drivers for change include the emergence of free content services (such as ResearchGate or SciHub) and increased popularity of publishing in open access and preprint repositories. It has become difficult for libraries to gauge the importance to researchers of database subscriptions. Reviewing usage statistics alone paints an incomplete picture and, as with all data, requires context from the users themselves.

Further complicating the picture are the results of a 2015 Ithaka S+R faculty survey at the University of Pittsburgh, which provided insight regarding the library's role in the discovery phase of the research life cycle. In the study, early-career researchers, especially those in the sciences, were the least likely population to agree that "the library serves as a starting point or 'gateway' for locating information for my

research." This finding, especially from the early-career science faculty, was concerning because it singled out a group of researchers in disciplines whose resources cost more than those in the social sciences or humanities. With a desire to better understand their behavior, we determined that the only course of action was to meet with this user group and ask them about their discovery process.

Why It Matters to Research Libraries

The converging contemporary issues of rising library subscription costs, a declining or uncertain budget outlook for library materials, and scholarly communication flowing outside of the traditional reach of the library make this the perfect time to assess what historical decisions we may want to rethink in order to create a different future. Akin to the canary in the coal mine, the authors of this project discuss the implications of early-career researchers going elsewhere for information and, more importantly, primarily relying on content that is not found through a library subscription. What can the present reality tell us about how searching behavior may look in 5 or 10 years? We needed to thoroughly investigate this question to assess what libraries should do to better meet the needs of the future research population of our institution.

Our original research question, "How does the library help to increase research productivity and impact?" with a focus on the discovery tools, is essential because of shifts in scholarly publishing, continued increases in the cost of A & I databases, and changing behavior of researchers. Additionally, if the library's role in the discovery phase of the research life cycle has changed, what opportunities does this present for libraries? How should libraries reposition themselves to meet the needs of early-career researchers?

Objectives

The objective of this study was to understand the information-seeking behavior of early-career faculty in hard science fields. We sought to learn how they discover published content and if they use library-purchased commercial databases in this process. Additionally, we wished to understand how much time and effort researchers invest when looking for content and what pain points they experience. Another objective of this project was to develop and evaluate a methodology for collecting and analyzing information about the library's role in the research discovery process. Our

hope was that this methodology could be extended to measuring the role of a library in other parts of the research life cycle. For instance, our partner, the University of Washington, applied the same methodology to ascertain the impact of libraries in the assessment/impact stage of the research life cycle.

The answers to these questions will help libraries understand how our next generation of researchers interact with libraries throughout the research life cycle. Thus, the results will guide library services, as well as shed light on ways libraries can support researchers in this process.

Hypothesis

Based on Pitt's 2015 Ithaka S+R faculty survey results,[3] we may assume that early-career faculty in STEM fields do not rely on publisher databases to discover information, because their level of agreement with the statement that a library is the "gateway to information" is very low. However, since we approached our research method using grounded theory, we refrained from formulating any formal hypothesis before analyzing the collected data.

Literature Review

A review of recent literature provided a foundation for our study.

Costs

Licensed resources are an increasingly heavy burden for libraries to bear, especially in the sciences. Librarians have been warily charting the steadily increasing costs of library collections for many years. The Association of Research Libraries created a startling visualization of the trajectory of serials costs.[4] In the last decade, the annual increases for subscription publications have been documented along with the disproportionately high prices of STEM subscriptions.[5] The focus on the cost of providing information to the academic community has reached the point where a press officer from George Washington University has made the following comparison: "Scholarly resources are not luxury goods. But they are priced as though they were."[6] We believe that many of the readers of this report will agree with this statement.

Discovery Process

While costs for abstracting and indexing databases continue to climb, the 2018 Ithaka S+R US Faculty Survey reports that only 33% of faculty in the sciences say they start their research with a specific scholarly database, down from 47% in 2015.[7] Aligned with these numbers is an increasing interest in examining the potential of freely available search engines like Google Scholar to not just supplement, but to replace A&I databases.[8] An interesting example of how the landscape has changed over the last decade can be seen in a study, first published in 2008 and replicated 10 years later, comparing search results in Google Scholar and Compendex, the latter considered a key source for universities supporting engineering. The 2018 paper concluded that, due to both an improved search engine and content, "Google Scholar is a reasonable alternative to expensive, fee-based tools."[9] And in the current environment, where Google and Google Scholar play an increasingly prominent role in discovery, it has been wryly noted that despite librarians' best efforts to steer students and faculty towards licensed library databases, users routinely began their search on Google Scholar.[10]

Though Google is a common starting point in the discovery process, an important aspect of research—a formal literature review—often still relies on academic search engines. These search engines, unlike Google, allow for transparency and reproducibility of search findings.[11] Increasingly, these tools are becoming available directly to researchers, free of charge (e.g., Dimensions or Meta) or via personal subscriptions (AlphaSense). They are also frequently built using AI tools to improve both recall and precision and to allow for natural language searching.

Google Scholar was launched in 2004, and the early-career researchers interviewed here, in most cases, began their PhDs by the mid-2000s. Graduate students, especially those in engineering, were early adopters of the Internet when starting the discovery process.[12] A longitudinal study on faculty reading behavior noted that, in general, faculty were spending less time reading articles, as well as reading fewer of them. It has been speculated that this change could be due to the amount of information that any person could reasonably consume, or it could be caused by the sheer amount of information so readily available to scholars via the traditional routes (journal alerts, books, etc.), as well as the creep of blogs, electronic mailing lists, and social networking sites into the academic research landscape.[13]

Early-Career Researchers

Although it is not only younger researchers who gravitate towards discovery systems with simple interfaces, an abundance of research on this cohort has been published from the Harbinger Project. This three-year research study, commissioned by the Publishing Research Consortium, succinctly points to "ubiquity, simplicity, and efficiency" as the reason so many early-career researchers start the research process with Google Scholar instead of library-subscribed discovery tools.[14] Paywalls are an additional constant pain point. As millennials, they are comfortable with the idea that "shadow libraries" like ResearchGate and SciHub are a reasonable place to visit after first trying the university library for the article they are trying to locate.[15] Even academic libraries with robust online access to journals are not enough of a pull to be the first place younger scholars go to locate articles.[16]

Methodology

The goal of our project was to collect insights to inform our services for researchers and identify new methods for evaluating the impact of these services. We wanted to create a manageable data collection and analysis model that can be extended to assess a library's role in various stages of a research life cycle. This project focused on the discovery stage.

To meet our goal, we employed elements of the grounded theory approach to data collection and analysis. This approach, first proposed by Glasser and Strauss, relies on structured and iterative analysis of qualitative data through categorization (or coding) and uncovering the connections between the emerging categories.[17] These categories and emergent theory are, therefore, "grounded" in the data. This approach does not use preexisting codes or categories, but, rather, uses the codes that emerge through a close reading of data.

Previous studies have examined how faculty across disciplines perceive and utilize the academic library throughout the research life cycle. Due to the highly individualized nature of this process, qualitative methods using interviews and focus groups have revealed complex data. Studies focusing primarily on quantitative analysis of this question have further identified the value of an interview-based approach and the gaps in understanding that this methodology will elucidate. Niu and Hemminger acknowledge that interviews are necessary to "better understand the underlying motivation of information-seeking behavior among scientists," and they expound on

trends in survey responses.[18] And Monroe-Gulick et al. indicated that "the open-ended nature of focus group discussions and interview interactions enabled participants in the current study to provide context and insight into their needs, highlighting nuances of their research requirements in ways the libraries could not have anticipated when constructing a strictly quantitative survey."[19]

In our study, we used semistructured interviews to collect data. The number of interviews for the sample was informed by Guest, Bunce, and Johnson's research demonstrating that data saturation can be achieved at the point of about 12 interviews.[20]

Semistructured interviews, as a data collection tool, allow for a deeper exploration of topics and emerging themes than a survey or a structured interview protocol would. The semistructured approach uses an interview guide with clearly articulated questions, but it also encourages an interviewer to stray from that guide to follow new leads or seek clarification. If done well, this method allows for establishing a rapport between an interviewer and an interviewee, encouraging them to be more forthright in describing their practices and articulating their perceived pain points. We felt this approach would provide us with more valuable data than, for instance, a direct observation or a survey approach.

In the process of developing our data collection tool, we met with Dr. Margarete Roller, a social sciences research consultant contracted by the Association of Research Libraries. Dr. Roller assisted the team with developing and sequencing the interview questions to best meet the goals of our project. The interview guide followed a "funnel" approach recommended by Dr. Roller and can be found in Appendix 1. It consisted of four parts:

1. Establishing rapport and building context
2. Delving into specific information-seeking behaviors
3. Overcoming challenges in finding research information
4. Describing an ideal state or a solution

The first set of questions was an opportunity for the researcher to talk about their field of study and their disciplinary norms for the dissemination of research outputs. This encouraged rapport and provided a friendly opening. Next, we sought to understand and quantify the ways that the researchers went about searching for information. For instance, what were their "go-to" resources, how much time were they prepared to spend on searching, and how or when did they know they had

"enough" information? The questions about challenges prompted the faculty member to make suggestions on ways the library can be more effective. The "ideal state" question allowed the researchers to reflect on their experiences and suggest ways to bring information-seeking to the next level.

Key Performance Indicators

This study was exploratory and designed to understand researcher behaviors rather than measure particular objectives through key performance indicators.

Methods

To collect data for this project, we set out to conduct a handful of in-person interviews with Pitt faculty. We focused our attention on the tenure-stream faculty in hard sciences and engineering, as a proxy for active research faculty. We defined early-career researchers as those who have completed their dissertations within the last 10 years. We specifically wanted to interview faculty active in research and those who were still fairly early in their academic careers, rather than those who were predominantly involved in teaching. To identify researchers who fit our definition, we conducted a preliminary search of the university faculty information system data and supplemented it with additional searches of departmental websites. This search yielded a list of 86 researchers. We reviewed their departmental affiliations and excluded faculty from the schools of the health sciences who are served by a separate library system.

The University of Pittsburgh institutional review board (IRB) approval for the project was obtained in January 2020. In early February 2020, we issued invitations to 77 researchers to participate in 60-minute in-person interviews, followed by another invitation a week later. In the end, 12 interviews were scheduled and conducted. Team members assigned to interviews conducted background research to orient themselves about research interests and current projects of the faculty they were about to interview. This information came from faculty web pages and recent publications. All interviews were completed by mid-March 2020—just a few days before the campus closed due to COVID-19. These interviews took place in person, in faculty offices, or in a library building, and were digitally recorded. We did not use a separate notetaker to assist during the interviews because we did not want to decrease the rapport between

the interviewer and the interviewee, and because of resource constraints. All interviews were transcribed by an IRB-approved commercial transcription service. Next, we prepared the transcripts for coding by reviewing the text for acronyms and ambiguous terms and removing any personally identifiable information.

The table below shows the departmental affiliation of the interviewees. The researchers interviewed represent a diverse cross section from the sciences and engineering.

Table 1: Departmental affiliation of research participants

Department	Number of Participants
Biological Sciences	3
Chemistry	1
Neurobiology	1
Mechanical and Material Science Engineering	3
Chemical and Petroleum Engineering	1
Informatics and Networked Systems	1
Industrial Engineering	1
Civil and Environmental Engineering	1

Once all of the transcripts were obtained, the project team discussed ways to approach coding and data analysis. It was agreed to loosely follow the themes from the interview questions and code the text based on these themes. Initially, all team members reviewed the same interview transcript independently; then we met to discuss the emerging themes and to agree on common vocabulary (codes). Next, each team member coded all remaining interviews independently and provided additional comments and observations (described as memos in the grounded theory approach). Team members then discussed their codes and notes to create an Excel table for capturing quotes and observations related to the emergent codes. This document was used for the subsequent description of the findings.

Resources Required

The team involved in this project comprised five members: a librarian with expertise in collection development in research libraries, two librarians with expertise in research and analytics, a liaison librarian for several science and engineering departments, and a library school intern.

The expertise from these individuals was critical in defining populations for the study, selecting methodologies for data collection, developing an interview guide, conducting the interviews, and analyzing the findings. Since few members of the team had prior experience in conducting interviews or working with qualitative data, customized sessions offered by ARL and Dr. Roller were invaluable in providing guidance on these aspects of the project.

The team recruited additional librarians to help with the interviews. We wanted to ensure that interviewers did not have any prior professional interactions with the respondents (e.g., as liaisons to their departments), to encourage the interviewees to respond with candor. One of the team members contacted potential interviewees, responded to them, and set up interview times. The library team used a portable recording device that some members of the group had used previously. Interviewers made sure they were facile in using this device. Once the interview was recorded, a staff member downloaded the file to a secure Box account. Next, a commercial company was identified to do transcriptions, which were generally completed within 48 hours. Then anonymized interview files and the transcribed files were made available in Box to the library team.

Findings

The interviews elicited far-reaching insights. For the purposes of review and analysis, we have grouped these responses into themes. Some of the themes were ones we deliberately sought out, but others emerged organically. The themes that emerged from the interviews encompassed: the nature of research, discovery and staying current, time and effort invested in discovery, diffusion of information sources, subscription content fulfillment, barriers and pain points, and opportunities for libraries.

Nature of Research

One observation we noted is that rather than being a one-off activity to prepare for a grant submission or a journal publication, the act of discovery is an ongoing process for these early-career faculty. Several interviewees noted that their research areas may shift based on funding availability. Searching for pertinent resources is a constant activity and spans several components of the research life cycle, including discovery, staying current, identifying adjacent/related research projects, identifying new collaborators, and identifying likely funding opportunities.

Another observation was that available funding can dictate the focus of these faculty members' research agendas. One interviewee commented bluntly that "science was a game about finding where the funding is, and finding sort of the question that the funding wants you to answer rather than following your passion or your interest or your idea." Another expressed a similar sentiment: "We diversify a lot as researchers now because of funding sources that you need to. You can't have a whole career in a single area."

We also noted the interdisciplinary nature of research interests, which necessitates reading widely and often outside a narrow disciplinary focus. One researcher expressed this idea by saying that "a lot of different disciplines are talking about them and having slightly different perspective[s] on how they approach them, but we can still learn a lot from each other. There's a lot of different literatures, in terms of journals that you look in and things like that, that add to making sure that you have a full picture."

Many research projects described by the interviewees were applied in nature, which often impacted the variety of information resources sought. Some examples of those resources were posters, meetings, personal websites of known researchers in the field, grant awards, and news of government and agency activities, as well as software tools and data storage platforms.

Because researchers in hard sciences and engineering tend to work in lab-based teams, the information discovery activity is often diffused among team members, often with graduate students conducting most of the searching and senior members mentoring their younger colleagues. An interviewee mentioned that "the students are the ones that are doing the more comprehensive lit review." On the other hand, some interviewees acknowledge that their students are savvier with new forms of technology. One mentioned, "I am not on Twitter, however, my technicians and programmers are and I learn a lot by what they tell me they saw on Twitter."

The faculty we spoke to feel the pressure to publish quickly and announce their research to avoid having their research scooped. They pointed out the time-consuming process of submitting and waiting for formal journal publication. One faculty member commented: "If they submit it to [a] journal, it takes around four to six months, and by the time the article was finished in the review process, another group published similar data and put it on arXiv."

Interviewees often also mentioned lack of time as a barrier to doing their work. One researcher noted that "the [discovery] process is important, and I wish I had double the amount of time in a day to do everything." Another said that "the biggest challenge to getting the information that I want is time. I don't have what feels like the room in my everyday schedule to keep up with the state-of-the-art in my own field."

Discovery and Staying Current

One major objective of this study was to identify what resources researchers use when they embark on discovery work. Researchers did mention a wide variety of resources that they use, and these are listed in Appendix 2. Though there are some commercial databases mentioned, over half of the resources referred to are freely available. Clearly, researchers have broad patterns of search activity and visit a wide array of sites. Often, these sites go beyond traditional academic research venues and include government, media, and social media sites, depending on their research focus.

All of the interviewees use Google or Google Scholar as their main discovery resource. They were all pleased with these sites and the features that Google makes available, such as cited references, related articles, and alerts. One researcher commented that "I can find everything from the most elementary knowledge, all the way to the most cutting edge stuff.... I find Google Scholar to be very convenient. It's fast and it's easy to use." Another said, "It's not probably the best approach but I usually will type in a search term first in Google Scholar and I'll just get a sense of what is out there." A third researcher pointed out "I just find it so much easier... I find it really intuitive."

Social media also plays a role in discovery. Twitter was mentioned by five of the interviewees. Again, all the researchers who used Twitter were pleased with this resource. They mentioned having to bypass the social posts, but the work posts of other researchers were extremely helpful to all. A faculty comment was: "I'm very active on Twitter, and I'd say that's probably the main source that I see new papers in my field." Interviewees mentioned the importance of identifying others working on related research and, thereby, expanding their own information net. Twitter was specifically mentioned as a primary resource for staying current: "I think that it's always a challenge to stay plugged into what's changing and what new things are out there. Twitter is an obvious one". Another researcher mentioned: "I really feel like it's an excellent place to make sure you don't miss anything." Facebook was mentioned but not seen as a valuable discovery tool.

Clarivate's Web of Science is the primary subscription resource for the scholars we interviewed. Researchers tend to use it when they are ready to submit a grant or other review paper. Though library-subscribed discovery tools were mentioned less often, faculty find the subscription tools that they do use to be important for validation and formalization of the discovery process. One researcher mentioned they find Google has more references but believed that Web of Science has a higher standard for including journals. Another researcher noted "the special use case for Web of Science for me is when you're not actually necessarily going to read every paper because you're either doing an automated content analysis type thing or whether you're trying to say the number of papers on this topic has gone up over time or something like that. You're not going to go through every one, and you want to trust that that search is returning what you actually think you're searching for. ...it's very, I would say, canonical, normal, traditional in our field to see someone say, 'We ran a Web of Science search for this query string.'"

Researchers have an intense interest in the discovery process. Several themes emerged in their responses to the interview questions. They feel that deep subject expertise is needed to uncover relevant information. They also see themselves as responsible for training their students in methods of searching for information while they see librarians in support roles.

Additional information resources were mentioned that are used in the critical activity of keeping up to date on activity in their field. For instance, most researchers mentioned attending meetings as a critical component in this process.

The researchers reported a variety of strategies for discovering relevant information, including employing alerts, performing themed searches using set keywords, looking for known researcher output, looking at the output of specific labs or research centers, and reviewing specific journal tables of contents. One interviewee said, "I use ISI alerts still, the Web of Science alerts. I have some that are targeted searches for certain keywords and I have some that just give me about six journals that I want to follow."

Time and Effort Invested in Discovery

One of our interview questions was to identify how much information researchers need to find and, relatedly, how far they will go to see/read an item that they have discovered but can't easily access. Interviewees seem to have interpreted this question with a focus on breadth of information.

There is a huge importance to these researchers in continually searching for what is current and new in their research space. In their efforts to keep up, faculty wish to discover all that is available. They would rather find as many resources as possible, rather than limiting the results of their searches. In response to the question "How do you know when you have enough information?" one scholar responded, "I never know. I don't think I ever have enough." This sentiment was echoed by another, who commented, "There's never enough…There's never a stop." Another respondent explained the process she used: "[You] try to extend your tendrils out from there, either looking through citations, doing reverse citation searches, figuring out not in the formal keyword search sense, but even just figuring out what phrasing they used in their abstract so that you know what to search for next. I guess the way I would usually do it is you just start sending those tendrils out until they start looping back on each other and at a certain point you're encountering 80% of the papers you've already seen and you say, 'Okay, that's probably about it.'"

Diffusion of Information Sources

The research outputs are incredibly varied, therefore the researchers acknowledged the need to consult a variety of types of sources and formats which are traditionally not covered in an academic database. For instance, in addition to journal articles, scientists cite data, software, grants, tweets, the Citizen Science Lab, videos, presentations, and patents.

The fragmented landscape of information sources makes it more difficult for librarians to assist researchers in the research process. As one respondent commented, "it's become a lot harder for the libraries to assist in the research process, the more specialized the research process has gotten. Because, although people who work in libraries are excellent, I wouldn't expect them to actually be able to help me do a lot of this. Because the knowledge that's necessary to know even what type of search to run is not something I would expect to be housed in the library anymore." This perception influences the relationship between researchers and librarians, potentially limiting the role of the library. The traditional role of the librarian recommending journal articles has diminished, and the researchers do not see the librarians as still playing a role in this new environment.

Subscription Content Fulfillment

In these researchers' views, the library plays a major role in fulfillment activities. Namely, researchers rely on the library as the source of the full-text content for materials they find, and they are often satisfied with the library's ability to provide these resources. "I mean, Pitt Libraries have access to everything I read regularly, but you'll end up in these obscure journals where it wasn't part of the bundle. I actually wouldn't consider that a major barrier."

But they note that if the material isn't available digitally it might as well not exist. One comment, "If it's not digital I'm not going to find it," refers both to the discovery process and an impatience with waiting for delivery, even though this content may be fulfilled through a colleague at another institution with a subscription.

Barriers and Pain Points

Several themes emerged about the research process in general when we asked researchers about barriers they encountered. In the discovery process, they identified a lack of good curation, that they often had to work to find valuable citations and could not rely on platform tools to winnow information. Faculty expressed several ideas for improvement, such as (1) creating a star system to highlight the best research

articles "so that people don't waste their time reading papers that don't really matter, but they can just scrape off the five-star stuff," (2) employing personalization metrics ["...it would be so much better if the information providers, they can sort out or, in a sense, predict what we want. For the very least, they can give us the most highly cited paper under a certain keyword"]; and (3) suggesting alternative terms "if there were something online that you could put in words and it gave you words in relevant fields." On the other hand, researchers indicated a willingness to dig deep to uncover the best sources. In the fulfillment stage, a barrier was having to wait more than 24 hours for full-text delivery.

Researchers complained about links not working, as well as about the number of click-throughs required to get to paywalled information. In reviewing the results of their searching, they find they still need to do a lot of work to unearth the best sources and tease out the validity of each piece of research. They are concerned that there is generally a low barrier to publication. One researcher summarized these sentiments: "I tend to see the body of research literature as having a finite, not insignificant but finite fraction of sort of gold nuggets of really meaningful information....I'm of the opinion that the barrier to publication has become low enough with the propagation of many, many, many journals that a lot of research results that get published are underbaked."

Researchers detailed pain points that involve the information landscape in several ways. They have a nagging concern that there is important work being done that is connected to their field of study, but that this work is being done in a neighboring discipline that uses different terminology and therefore is hidden from their discovery methods. They also mentioned the uncertainty about keywords and normative terms when they tried to move out of their area of expertise and into multidisciplinary fields. One faculty member mentioned, "When talking to other people, it seems that different disciplines will use different terminology to talk about the same ideas," while another said, "Usually if I can't find enough information, it's because I don't have the language or the terminology to access it and find it."

Opportunities for Libraries

Interviewees had a variety of takeaways related to library services. They see the library as the campus provider of content. Several interviewees mentioned librarians having a role in teaching data skills and providing bibliographic instruction for their students. As was noted above, though, some faculty question whether librarians have the disciplinary expertise to do the exhaustive database searching required.

Interviewees mentioned several other potential library services that could be helpful in supporting their research, including repositories of good review papers and topic overviews, systems that predict which other papers would be helpful to them, better linking of results sets by related topics, and tools that translate the language of one field into the language of another. Researchers expressed a need for enhanced discovery tools using better algorithms and AI to develop more targeted retrieval.

Some researchers acknowledged the value of libraries as archives of research output in all formats: "The archival practices play some role there because, in theory, you [the library] live longer than any lab does. You [the library], even in theory, live longer than any department does. Whole fields of study should be able to come and go and the library is still there, right? I don't think that role has gone away. It's just harder because there's more stuff."

The library's role in content management is being reenvisioned: no longer the steward of a unified collection, the library becomes the facilitator of a networked suite of open and extensible tools, resources, and services.[21]

Data

The data collected in this project was qualitative and consisted of transcripts of semistructured, in-person interviews conducted in February–March 2020. In the process, we collected nearly 12 hours of recordings, which translated into 207 pages of transcripts. The transcripts were coded and resulted in 60 Excel worksheets, which were then distilled into a single document of common themes and observations.

The cost of collecting the data included both the time of the team members in designing the interview guide and the time to conduct the interviews (each was scheduled to last up to 60 minutes), as well as the time to code and analyze the transcripts. We opted to use a commercial transcription service, which required a payment of $1.25 per minute of transcription. The team used two digital recorders available from the ULS and password-protected cloud storage for files already available to us via a Pitt subscription.

The primary obstacle to gathering and analyzing the data included the large time commitment involved. To overcome any scheduling conflicts, we recruited additional interviewers to assist us in data collection. The coding of the transcripts was also very time-consuming, but we decided that all investigators would code all of the transcripts because this investment of time upfront would assist us with the data analysis later.

Value

Overall, the project fulfilled the objectives of our team in providing a robust understanding of early-career faculty and postdoctoral research needs in the area of discovery.

We decided to use in-depth semistructured interviews to collect data for our project. This approach proved to be time-consuming in both the data collection and analysis stages. It required additional training from team members in working with qualitative data. On the other hand, we felt that we collected more nuanced information from our respondents than a questionnaire, often used in library research, would allow for. We felt that through this "conversation" we gained a better understanding of the process and environmental challenges from the researchers' perspectives, which, in turn, provided us with new perspectives on how libraries can fit within that process. For instance, the near-constant search for funding and the increased interdisciplinarity, on the one hand, and specialization, on the other, of research problems, and as well as the responsibility for training junior researchers, all create demands on researchers' time. These findings led us to conversations about our current educational outreach to graduate students and about the role lab leaders may play in helping us more closely align our efforts with the needs of their teams. We also noted the opportunities in sharing our findings with other research support services on campus to reflect on our efforts and to consider providing more integrated approaches.

Our findings related closely to the information discovery process itself and confirmed that the role of subscription-based A&I databases has been superseded by services such as Google Scholar and some social media platforms. Interestingly, faculty noted that including the evidence of a literature search from sources such as Web of Science is still required in some disciplines (e.g., to be included in grant proposals or publications). Seamless and timely access to full-text materials, though, is critical, and researchers rely on institutional subscriptions for access. However, when institutional access is not available, they are more likely to use their own networks of contacts (including SciHub) to access the full text rather than rely on the library's interlibrary

loan service, which they perceive as too slow. These findings will allow the library to redirect our resources from some A&I services in order to better support full-text access. Our findings also raise several questions relating to possible improvements for more seamless paths to access and for reviewing and supplementing our resource sharing services.

Apart from the findings themselves, we also benefited in other ways by participating in this research, including the following:

- **Engaging a new generation of Pitt researchers**
  Another important aspect of this project was an opportunity to speak directly to a subset of Pitt's research community. Increasingly, our engagement with this group was limited to securing access to requested journal subscriptions. Thus, their perception of library value to them was also limited to that one aspect. Through these interviews, we opened up potential new avenues for engagement, particularly related to graduate student training and support. Even though researchers see themselves as ultimately responsible for students' training, they also recognize that the library can play a role in the process and might be open to working with us on customizing such training.

- **Acquisition of new skills by team members**
  For many of us, developing a survey tool, conducting interviews, and coding and analyzing collected qualitative data was a new experience. The training opportunities provided by ARL were invaluable in this process. Internally, we gained a better understanding of the IRB process, built a cross-departmental team, and engaged additional colleagues to assist in parts of the process, thus strengthening institutional research capabilities. Indeed, some members of the team have now successfully participated in another ARL project requiring a similar skill set.

- **Partnering with the University of Washington team**
  Having a research partner was very important in our ability to frame the project in a broader context of the research life cycle. While each institution focused on different aspects of library support for research, the framework allowed us to think about our projects in a more cohesive manner, allowing us to use similar methodologies and tools for data collection. This, in turn, allowed for collaborative work on the development and validation of these tools. Moreover, having an external partner helped keep us on track and adhere to deadlines through monthly meetings. These meetings also proved to be a great

forum for discussion of approaches to data analysis, initial findings, and development of research outputs (e.g., a joint Library Assessment Conference presentation and paper).

- **Value to other libraries**
  We believe that findings from the project, as well as the survey instrument developed to collect data, can be of interest to other research libraries in their quest to understand the information discovery process of early-career researchers in science and technology.

Lessons Learned

We felt that the research and process were worthwhile. We have findings to act on, and we have a data collection tool and an established process that we can use in the future with different groups of researchers. Other opportunities for further research include supplementing qualitative data collected through interviews with quantitative data from proxy logs or COUNTER and other usage reports.

The interview process and the use of semistructured interview questions worked well. The faculty members were positive about signing up for the interviews and saw it as part of their responsibility to assist librarian research. The interviews were nonthreatening and comfortable, and the faculty members were open and conversational. They shared details about their work and seemed to speak honestly about the processes that work for them, as well as the ones that do not. A fairly small number of interviews were conducted, and the resulting transcription and analysis were straightforward. We are pleased with the process and we feel that we gained valuable information about our users' behavior. This information can provide direction for decisions regarding collection development and services.

Some areas for improvement include:

- More training for interviewers in developing skills needed to encourage candor and rapport with interviewees. For most of the interviewers, this was their first experience conducting interviews of this type. We could consider more training using mock interviews and reviewing examples of "successful" interviews.

- Consider using software for coding qualitative data for a more streamlined (less time-consuming?) process.
- Improve engagement with ARL throughout the process, including better alignment between ARL and project teams' goals in the initial proposal stage.

Recommendations for Future Research

The role of the library in the research process is at a crossroads. The information gleaned from the early-career researchers who provided feedback by taking the Ithaka S+R survey—first in 2015 and then again in 2018—has signaled a change in user behavior that we need to continue to monitor. The conversations captured during these interviews, too, have allowed us a tremendous opportunity to get a sample of how tenure track faculty in engineering and science approach their work. Most importantly, the immediacy of these interviews offers a chance to take a breath to consider what other systems we have developed to support the work of *all* research faculty. Within the last 10 years, the ULS reorganized in order to focus on helping faculty navigate the complexity of the various aspects of scholarship in the 21st century, including scholarly communication, data management, a university repository, and ORCiD adoption. We also partnered with university IT to implement and provide outreach to a faculty information system. The results from this work will allow us to step back to study the changing landscape and ask ourselves how we can change what *we* offer beyond providing an alternative to Google Scholar.

Recommended Reading

American Library Association. "Prices of US and Foreign Published Materials." Accessed February 5, 2021. http://www.ala.org/alcts/resources/collect/serials/spi.

Chen, Xiaotian. "The Declining Value of Subscription-based Abstracting and Indexing Services in the New Knowledge Dissemination Era," *Serials Review* 36, no. 2 (December 6, 2013): 79–85. DOI: https://www.tandfonline.com/doi/abs/10.1080/00987913.2010.10765288.

Cooper, Danielle, Rebecca Springer, Jessica G. Benner, David Bloom, Erin Carrillo, Alexander Carroll, Bertha Chang, et al. *Supporting the Changing Research Practices of Civil and Environmental Engineering Scholars* (Ithaka S+R, January 16, 2019). https://doi.org/10.18665/sr.310885.

Engel, Debra, Sarah Robbins, and Christina Kulp. "The Information-Seeking Habits of Engineering Faculty." *College and Research Libraries* 72, no. 6 (November 2011): 548–567. https://doi.org/10.5860/crl-155.

Gardner, Tracy, and Simon Inger. *How Readers Discover Content in Scholarly Publications* (Renew Publishing Consultants, August 2018). https://renewconsultants.com/wp-content/uploads/2018/08/How-Readers-Discover-Content-2018-Published-180903.pdf.

Wellings, Susan, and Biddy Casselden. "An Exploration into the Information-Seeking Behaviours of Engineers and Scientists." *Journal of Librarianship and Information Science* 51, no. 3 (September 1, 2019): 789–800. https://doi.org/10.1177/0961000617742466.

Zhang, Li. "Use of Library Services by Engineering Faculty at Mississippi State University, a Large Land Grant Institution." *Science & Technology Libraries* 34, no. 3 (November 7, 2015): 272–286. https://doi.org/10.1080/0194262X.2015.1090941.

Zoellner, Kate, Samantha Hines, Teressa Keenan, and Sue Samson. "Faculty Research and Publication Practices." *portal: Libraries and the Academy* 15, no. 1 (January 2015): 111–131. https://doi.org/10.1353/pla.2015.0009.

Appendix 1: Interview Guide

University of Pittsburgh University Library System

In-depth interviews with early-career faculty concerning discovery of published content

1. Building Context: Current Research Area of Focus [Note: This provides the foundation and context for the rest of the interview, and helps establish a conversational tone and build rapport.]
    a. Broadly speaking, describe your research area of interest.
        i. What is your primary research question or objective?
        ii. How long have you been researching this area?
        iii. What got you interested in this research area (or topic or issue)?
    b. Describe for me the current project(s) you are working on to address this research objective.
        i. What do you hope will be the outcomes or "products" from this/these projects? (What are the artifacts and how are they communicated?)
    c. Is there a lot going on in this research area right now?
        i. How much is published or written about on this topic or issue? (Is this a new, breakthrough, or established area?)
        ii. How do you keep up with what is going on in the field? (Where do they look—print sources, people, conferences, routines, other?)

2. Information Seeking — Attitudes & Behavior
    a. Through your research process you may need to find information. For instance, How do you typically go about doing this? [Note: Listen for activities related to gaining information.]
        i. What type of information are you looking for/do you need?
        ii. What is the best source for this information?
            1. Where/how do you access this source? [NOTE: Listen for library and probe]
            2. What makes this the best source? [Possible responses: easy/fast, convenient, free, most complete. Probe to clarify meaning for participant.]

b. What is the worst source for this information?
    i. What makes this the worst source?
c. I noticed you have not mentioned X. Do you ever use it (for instance, in teaching, research, promotion, other)? [Probe as required: Google Scholar, Research Gate, SciHub, SCOPUS, WOS, various databases (probe for which and how they are accessed).]
d. How do you know when you have what you need for your research purposes? (How much is enough?)
    i. What do you do when you cannot find "enough" information?

3. Overcoming Challenges & Opportunities for the Library
    a. You have talked about the best and the worst sources for the discovery of information you need. How would you describe your biggest challenge to gaining the information you need?
        i. What would remove or ease these challenges?
    b. What, if any, role could the library play in easing these challenges or facilitating your information gathering?

4. "Ideal" Information Gathering
    a. Before I let you go, I have one final question. We have talked a lot today about your research process and, specifically, how you gain information to inform your research. If you could create an "ideal" scenario of the information-gathering process for you and your work, what would it be? There are no rules here. I am simply asking you to imagine an "ideal" way to discover information.

Appendix 2: Table of the Discovery Resources Mentioned

Resource	Number of times mentioned	Department affiliation of interviewees
Google/Google Scholar	12	All
Web of Science	7	Biological Sciences, Chemical and Petroleum Engineering, Mechanical and Materials Science Engineering, Civil and Environmental Engineering, Chemistry
Twitter	5	Biological Sciences, Chemical and Petroleum Engineering, Neurobiology
Scopus	5	Biological Sciences, Mechanical and Materials Science Engineering, Civil and Environmental Engineering, Industrial Engineering
PittCat/library website	3	Biological Sciences, Informatics and Networked Systems, Industrial Engineering
PubMed	3	Biological Sciences, Neurobiology, Informatics and Networked Systems

Resource	Number of times mentioned	Department affiliation of interviewees
Scifinder Scholar	3	Chemical and Petroleum Engineering, Mechanical and Materials Science Engineering, Civil and Environmental Engineering
Compendex	2	Civil and Environmental Engineering, Industrial Engineering
Dryad/Figshare	2	Biological Sciences
Engineering Village	2	Mechanical and Materials Science Engineering, Industrial Engineering
Github	2	Biological Sciences, Mechanical and Materials Science Engineering
ResearchGate	2	Chemical and Petroleum Engineering
Personal websites	1	Biological Sciences
bioRxiv	1	Neurobiology
Government websites	1	Mechanical and Materials Science Engineering

Resource	Number of times mentioned	Department affiliation of interviewees
BCC database	1	Civil and Environmental Engineering
Proceedings	1	Mechanical and Materials Science Engineering
Theses	1	Biological Sciences
Knovel	1	Biological Sciences
Facebook	1	Biological Sciences

Appendix 3: Themes and Codes

Department, research interests

Outputs of research

Keeping up

Process to find info

Best sources

Worst sources

Other comments finding

How much is enough

Challenges

Solutions to challenges

Role of library

Ideal process

Endnotes

[1] Anam Mian and Gary Roebuck, *ARL Statistics 2018–2019* (Washington, DC: Association of Research Libraries, 2020).
[2] Anam Mian and Gary Roebuck, *ARL Statistics 2018–2019* (Washington, DC: Association of Research Libraries, 2020), https://doi.org/10.29242/stats.2018-2019.
[3] Ithaka S+R Faculty Survey: results for University of Pittsburgh, 2015, unpublished.
[4] Martha Kyrillidou and Mark Young, comps., *ARL Statistics 2003–04: A Compilation of Statistics from the One Hundred and Twenty-Three Members of the Association of Research Libraries* (Washington, DC: Association of Research Libraries, 2005), 10–24, https://files.eric.ed.gov/fulltext/ED498294.pdf.
[5] Stephen Bosch, Barbara Albee, and Sion Romaine, "Follow the Money," *Library Journal* 145, no. 4 (2020): 36–41; Kyrillidou and Young, *ARL Statistics 2003–04;* "EBSCO Information Services Releases Serials Price Projection Report for 2019," EBSCO, September 20, 2018, https://www.ebsco.com/news-center/press-releases/ebsco-information-services-releases-serials-price-projection-report-2019.
[6] George Washington University Libraries,*Trying Times: The Impact of Inflation on the GW Libraries Budget*, https://perma.cc/55AK-TCFL.
[7] Melissa Blankstein and Christine Wolff-Eisenberg, *Ithaka S+R US Faculty Survey 2018* (Ithaka S+R, April 12, 2019), https://doi.org/10.18665/sr.311199.
[8] Lotta Haglund and Per Olsson, "The Impact on University Libraries of Changes in Information Behavior among Academic Researchers: A Multiple Case Study," *The Journal of Academic Librarianship* 34, no. 1 (January 2008): 52–59, https://doi.org/10.1016/j.acalib.2007.11.010; John J. Meier and Thomas W. Conkling, "Google Scholar's Coverage of the Engineering Literature: An Empirical Study," *The Journal of Academic Librarianship* 34, no. 3 (May 2008): 196–201, https://doi.org/10.1016/j.acalib.2008.03.002.
[9] Carmen Cole et al., "Google Scholar's Coverage of the Engineering Literature 10 Years Later," *The Journal of Academic Librarianship* 44, no. 3 (May 2018): 419–425, https://doi.org/10.1016/j.acalib.2018.02.013.
[10] Jeremy Cusker, "Elsevier Compendex and Google Scholar: A Quantitative Comparison of Two Resources for Engineering Research and an Update to Prior Comparisons," *The Journal of Academic Librarianship* 39, no. 3 (May 2013): 241–243, https://doi.org/10.1016/j.acalib.2013.02.001.
[11] Andrea Polonioli, "In Search of Better Science: On the Epistemic Costs of Systematic Reviews and the Need for a Pluralistic Stance to Literature Search," *Scientometrics* 122 (2020): 1267–1274, https://doi.org/10.1007/s11192-019-03333-3.
[12] Carol George et al., "Scholarly Use of Information: Graduate Students' Information Seeking Behaviour," *Information Research* 11, no. 4 (July 2006), http://InformationR.net/ir/11-4/paper272.html.
[13] Carol Tenopir et al., "Scholarly Article Seeking, Reading, and Use: A Continuing Evolution from Print to Electronic in the Sciences and Social Sciences," *Learned Publishing* 28, no. 2 (April 2015): 93–105, https://doi.org/10.1087/20150203.

[14] David Nicholas et al., "Does the Scholarly Communication System Satisfy the Beliefs and Aspirations of New Researchers? Summarizing the Harbingers Research," *Learned Publishing* 33, no. 2 (April 2020): 132–141, https://doi.org/10.1002/leap.1284.
[15] Jeremy S. Faust, "Sci-Hub: A Solution to the Problem of Paywalls, or Merely a Diagnosis of a Broken System?" *Annals of Emergency Medicine* 68, no. 1 (July 1, 2016): A15–A17, https://doi.org/10.1016/j.annemergmed.2016.05.010; John Bohannon, "Who's Downloading Pirated Papers? Everyone," *Science* 352, no. 6285 (April 29, 2016): 508–13, https://doi.org/10.1126/science.352.6285.508.
[16] Nicholas, "Does the Scholarly Communication System Satisfy the Beliefs and Aspirations of New Researchers?"; Faust, "Sci-Hub: A Solution to the Problem of Paywalls, or Merely a Diagnosis of a Broken System?"
[17] Barney S. Glasser and Anselm L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research* (Chicago: Aldine, 1967).
[18] Xi Niu and Bradley M. Hemminger, "A Study of Factors That Affect the Information-Seeking Behavior of Academic Scientists" *Journal of the American Society for Information Science and Technology* 63, no. 2 (February 2012): 336–353, https://doi.org/10.1002/asi.21669.
[19] Amalia Monroe-Gulick, Greta Valentine, and Jamene Brooks-Kieffer, "'You Need to Have a Street Beat': A Qualitative Study of Faculty Research Needs and Challenges," *portal: Libraries and the Academy* 17, no. 4 (October 2017): 777–802, https://doi.org/10.1353/pla.2017.0046.
[20] Greg Guest, Arwen Bunce, and Laura Johnson, "How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability," *Field Methods* 18, no. 1 (February 1, 2006): 59–82, https://doi.org/10.1177/1525822X05279903.
[21] Sarah Lippincott, *Mapping the Current Landscape of Research Library Engagement with Emerging Technologies in Research and Learning*, edited by Mary Lee Kennedy, Clifford Lynch, and Scout Calvert, Association of Research Libraries, Born-Digital, Coalition for Networked Information, and EDUCAUSE, 2021, https://doi.org/10.29242/report.emergingtech2020.landscape.