



Personal Library Curation: An Ethnographic Study of Scholars' Information Practices

Smiljana Antonijević and Ellysa Stern Cahoy

abstract: This paper presents findings of a Mellon Foundation-funded study conducted at Penn State University in University Park during Fall 2012 that explored scholars' information practices across disciplines encompassing the sciences, humanities, and social sciences. Drawing on results of the Web-based survey and ethnographic interviews, we present findings related to academics' practices in discovering, storing, citing, and archiving information, as well as their views on the role of library in the digital age. The paper harvests a comparative multidisciplinary perspective of our study, identifying principles and technical architecture that support digital scholarship and facilitate the development of literacies for faculty personal information management.

Introduction

As early as 1983, Jo Bell Whitlatch wrote, "Speculation concerning the future of libraries has often been focused upon the advent of the 'paperless society' and the effects of library automation."¹ Today, the physically based library as a repository of books and other materials is rapidly becoming a cultural anachronism. Perhaps nowhere is this more evident than within the academic library, which is transforming into an online entity as faculty and students find much of their scholarly information in electronic journals and other digital research resources.² Indeed, the most significant libraries in scholars' universe of today may be those that they are building themselves, on their own computers.

Today, the physically based library as a repository of books and other materials is rapidly becoming a cultural anachronism.

portal: Libraries and the Academy, Vol. 14, No. 2 (2014), pp. 279–298.

Copyright © 2014 by Johns Hopkins University Press, Baltimore, MD 21218.



This shift from institutional stewardship of information to individualized information management mandates a change in how librarians partner with users toward a new model for helping them effectively manage their personal information collections. Providing support for different phases of scholars' workflow—from finding, storing, citing, annotating, and sharing to archiving information—is a unique service challenge for libraries. For instance, David Seaman's study conducted among the humanities faculty indicated a need for library services to connect various phases of the workflow, such as "mass data storage, persistent identifiers, interlinked scholarly and pedagogical repositories, collaborative online work, community tagging, and user commentary."³ Scholars' individualized information management also brings to the forefront the issues of curation and archiving, as pointed out by Catherine Marshall, Sara Bly, and Françoise Brun-Cottan in 2007; by Dorothea Salo in 2009; and in the recent faculty survey by the nonprofit research and consulting group Ithaka S + R reported by Roger Schenfeld and Ross Housewright in 2013.⁴ Academics' practices of finding information have also transformed significantly in the digital age, as already observed twenty years ago by Susan Siegfried, Marcia Bates, and Deborah Wilde. Numerous studies afterward confirmed that observation, including those by Lokman Meho and Helen Tibbo in 2003; Allen Foster in 2004; Bradley Hemminger, Dihui Lu, K. T. L. Vaughan, and Stephanie J. Adams in 2007; Carole Palmer, Lauren Tefteau, and Carrie Pirmann in 2009; and Ellen Collins, Monica E. Bulger, and Eric T. Meyer in 2011.⁵

For all those reasons, studying scholars' digital workflow, including areas of ease and challenge, is integral to developing services that focus on the user as his or her own empowered librarian and steward of electronic collections. Such an approach was the core of our Andrew W. Mellon Foundation-funded project completed in 2013 at Penn State University and presented in this paper. In the project, we explored the digital workflow of the Penn State faculty across disciplines encompassing the sciences, humanities, and social sciences. The project examined scholars' practices and needs related to information management at all stages of the research cycle, as well as their broader attitudes toward digital scholarship. In this paper, we focus on project findings related to information management and retrieval, including finding, storing, citing, and archiving information, as well as on our respondents' attitudes toward the role of the library in the digital age. We present the results of a Web-based survey completed by 196 Penn State faculty members during Fall 2012, as well as the qualitative analysis of twenty-three ethnographic interviews conducted with Penn State faculty during the same period.

This paper harvests a comparative multidisciplinary perspective of our study to explore specificities of scholars' information management and retrieval practices across the academic disciplines, providing a ground to identify and develop a software and service architecture that supports those practices. Therefore, while focusing on current findings, the paper briefly highlights the future trajectory of our study, as well as planned next steps regarding technological initiatives aimed at addressing management of scholarly workflow.

In the following section, we describe the methodological approach used in our study, as well as the study participants' demographic information. We then turn to presenting our results, followed by the discussion and conclusion sections.

Methodology

Our study comprised two research phases, each of which focused on a specific set of research questions and goals. The first phase included a Web-based survey posted via a Web-based survey management platform called LimeSurvey (<http://www.limesurvey.org/>) in October 2012. The survey was available to Penn State faculty over two weeks. It consisted of twenty-five questions, which, in addition to demographic information, included queries about data searching, storing, citing, sharing, and archiving practices, as well as about scholars' experiences in using digital research tools and resources.

A total of 196 faculty respondents from across the sciences, social sciences, and humanities completed the survey; 58 percent were female, and 42 percent were male. A majority of respondents were tenured faculty, with tenure-track faculty and fixed-term (non-tenure track) faculty following. Overall, the humanities tended to have older respondents (over 40 years of age), while the sciences and social sciences skewed lower in age. Gender distribution of respondents remained static across the sciences and humanities, while a higher number of women in the social sciences participated in the survey.

Survey responses were collected using LimeSurvey. Unless respondents volunteered their contact information, all survey responses were anonymous, with only demographic data captured to classify individual respondents. The data were downloaded into Microsoft Excel 2010 (Excel) using LimeSurvey's data export feature. The data mining and cleansing stages were completed using Excel before being uploaded into IBM SPSS Statistics 20 (SPSS), a software package for statistical analysis. All statistical tests, tables, and descriptive statistics were created using SPSS.

The survey was intended as an investigative tool that enabled us to develop an aggregated insight into currently existing research practices of the Penn State faculty across colleges, academic disciplines, and faculty ranks. The survey also served as a recruitment tool for the second phase of our study, which included a set of face-to-face ethnographic interviews.

The interviews were designed to further explore patterns of faculty digital workflow, providing in-depth understanding of the examined practices, thus fine-tuning insights gathered in the first research phase. As previously mentioned, the interviewees were recruited from the poll of faculty members who took part in the Web-based survey. A total of twenty-three scholars volunteered to participate in the interviews. The interviewees were equally divided along the lines of disciplinary profiles, academic ranks, and gender: 13 were faculty in the humanities and social sciences and 10 in the sciences; 11 were tenure-track and 12 tenured faculty; 13 were female and 10 male.

The interviews were semi-structured and, on average, lasted an hour. Most of the interviews were conducted in the respondents' university offices.⁶ The interviewees were audio-recorded and then transcribed by a professional transcriptionist. The interview transcripts were first coded into broader categories (nodes) by two independent coders. We then proceeded with focused coding, where the categories into which the data were originally coded had additionally been refined for relevant patterns, themes, and topics. The interview results are presented anonymously, and the interview quotations have been edited for clarity and length.



The survey part of the study focused on information use and retrieval-centered aspects of the workflow, while the interviews had a broader scope. In this paper, we present results covered by both research methods; our other findings will be presented in a separate publication. Specifically, in the following section we are presenting our survey and interview results related to finding, storing, organizing, citing, and archiving information, as well as the study participant's attitudes toward the role of libraries in the digital age.

Results

Finding Information

In the workflow of a digital scholar, various information resources serve as search and access points, including library collections, disciplinary databases, and Web search engines, as well as scholars' personal digital libraries and other individually created resources. Among our respondents, the survey results indicate high placement of academic databases such as ProQuest, JSTOR, and others in academics' information searches (70 percent), followed by use of Google Scholar (63 percent), Google Search (57 percent), and the local Penn State Libraries catalog (44 percent). Resources not listed in the survey comprised 31 percent of cases; for instance, a number of faculty indicated footnote and bibliography analysis of seminal articles as a useful discovery method.

The interview results show that searching for information electronically is a standard, daily practice of our respondents, regardless of their disciplinary background or technical proficiency, and they consider it one of the most important benefits of

Searching for information electronically is a standard, daily practice of our respondents, regardless of their disciplinary background or technical proficiency.

incorporating digital technologies into their research practice. "If I'm looking for something I read 10 years ago and I remember a couple of keywords, I can usually find it much quicker than I could the old-fashioned way; that's fantastic. So, in terms of research, that's the big innovation," explains one of our respondents, a professor of chemistry.

The interview results also indicate that the path toward finding information commonly starts with Google Search and Google Scholar, especially for scholars engaged

The path toward finding information commonly starts with Google Search and Google Scholar, especially for scholars engaged in discovery search, the first round of academic research.

in discovery search, the first round of academic research. "If I'm starting out looking at a new topic, I usually start on the Web and then might go to the library's databases and look at more specific things," explains an assistant professor of English. Sometimes,

electronic library catalogs are also used for discovery search; “I typically go to the library’s main Web site and enter a search term through that means,” says an assistant professor of education, specifying that for him this type of search is exploratory in nature. The same professor reports that his other common search strategy is to search within the citation management software Papers, calling it a “more targeted effort [when] I know what I want and I simply want to get it and put it into a place that I can manage it.” Our findings about the significance of Google or Google Scholar as access points in searching for information reassert the results of some previous studies on information-seeking behavior, such as that by Nicholas and his coauthors.⁷ Academics also commonly use Web search engines, especially Google Scholar, to identify different connective structures, which include citations, specialized vocabularies, and so on. “I typically use Google Scholar, and I’ll work forward and backwards on that. I will find a paper that I know is relevant, but maybe written a few years back, see who cited it recently, then look at who they cited and start that path,” explains an assistant professor of engineering design.

Use of academic databases is also common among our respondents. For instance, an assistant professor of mathematics identifies arXiv—an archive for preprints of scientific papers in mathematics, physics, and related fields—and MathSciNet—the American Mathematical Society’s database of math articles—as his key information resources. An assistant professor of French explains that although she “continue[s] to use Google . . . for more specialized articles, [she] typically work[s] with JSTOR and Project MUSE,” referring to two databases of scholarly journals.

Library databases usually serve as an access point for scholars in the humanities, especially if they are engaged in a known item search. “What I normally do is I simply go to the MLA [Modern Language Association] bibliography and either download PDFs or order things through interlibrary loan,” says a professor of English. In general, however, discovery search tends to be increasingly related to commercial Web search engines rather than to academic databases. For example, a professor of physics explains that Web of Science—a commercial service that indexes journals

Library databases usually serve as an access point for scholars in the humanities, especially if they are engaged in a known item search.

in the sciences, social sciences, and humanities—provides easy access if the reference is previously known, but it “is not great for finding things that you don’t know exist.” The same experience of challenges in using library databases for information search is echoed in the words of our other respondents. “Sometimes the [library] database just doesn’t pick up things or you have to go to multiple databases, so then you’re going back and forth, and that’s not fun,” reports an associate professor of education. In the same manner, an assistant professor of education explains that, compared to the library online catalog, Google is “just quicker because you don’t have to go through so many steps . . . it’s just faster, there’s not so much clicking on so many different windows to get there.”

This finding reaffirms results of other recent studies indicating the increasing prevalence of commercial over academic services for scholars’ information search. For instance, Utrecht University Library in the Netherlands has recently completed a study among its users, and, after finding the prevailing use of commercial services for information search, decided to close down its own local discovery tool.⁸

Scholars' personal digital collections also serve as an important information resource. "All those PDFs that I've scanned years ago are here," explains an assistant professor of comparative literature. He details further how such availability influences his search practice: "Because everything's portable now, it's providing a function, an almost organic memory. I tend not to remember the things that I used to force myself to remember because it's just easier to pull it up and [it's also] more accurate that way."

Finally, a specific type of finding information is related to the use of digital information aggregators, that is, local and Web-based applications that tap multiple sources of data. Among our respondents, scholars in the sciences make more extensive use of such aggregators to stay abreast with latest developments in their field. "I receive weekly or monthly updates from a number of different aggregator services that summarize recent submissions or recent publications in areas of research that interest me," says a professor of astrophysics. For a professor of chemistry, the use of such aggregators is a daily routine:

Much of what I read comes from the American Chemical Society, the Royal Chemical Society, and through those applications the societies distribute the latest abstracts with figures, doing that on a daily basis, until late night hours. And so when I get into bed, everything for the day has been published and I can go through it.

The diversity of institutional and personal digital information sources that contemporary scholars encounter confronts them with some important challenges, such as information overload and difficulties in organizing materials, as we will describe in more detail in the next subsection.

Storing and Organizing Information

The part of scholarly workflow closely related to finding and gathering information is storing and organizing those materials in useful and reliable ways. The survey results indicate that most of our respondents actively store materials important to them. However,

Humanities and social sciences faculty are more likely to store Word documents, while faculty in the sciences are more likely to store data files and presentation files.

what faculty store differs across disciplines. For instance, humanities and social sciences faculty are more likely to store Word documents, while faculty in the sciences are more likely to store data files and presentation files. In both of these academic groups, the vast majority of respondents

(97 percent) reported storing PDFs. Furthermore, while e-mail represents a significant information collection, less than 60 percent of humanities and social sciences faculty, and just over 50 percent of faculty in the sciences, report intentionally storing it. This low response rate for e-mail storage may simply point to a low faculty understanding of automatic storage and backup provided by e-mail clients and other services.

Survey respondents overwhelmingly indicated that they most frequently store information on the hard drive of their computer or computers—95 percent of humanities and social sciences faculty and 90 percent of faculty in the sciences (92 percent of faculty

overall). The file-sharing service Dropbox was the second most often used storage solution, indicated by 50 percent of the surveyed faculty. Usage of any other different type of storage service was much lower across disciplines.

Our interview results indicate that the activity of storing and organizing information is not a straightforward task, for various reasons. To start with, our respondents' clearly distinguish between storing and organizing research articles and research data. And while the former constitutes less of a challenge, the latter stands as a difficulty across disciplines.

Our interviewees indicated that they commonly store and organize research articles in an electronic format, often through the use of research or citation management software; interviewees in the sciences commonly mentioned Papers or JabRef, while respondents in the humanities and social sciences identified EndNote and Zotero as popular bibliographic tools. "Papers, I have found to be exceptionally useful and handy for storing electronic manuscripts," says a professor of biology. He uses the same program for organizing research articles, explaining it in the following way: "When I've identified things that are useful, it [Papers] has a bunch of Smart Folders. And those papers automatically end up on the iPad. I make sure everything I read I give some sort of rating, and that's how I keep track of what I've read. That's by and large how I handle almost everything."

In the humanities and social sciences, EndNote and Zotero tend to be well-liked as citation managers, but they are not considered advantageous in terms of organizing materials. For instance, an assistant professor of comparative literature observes:

Zotero does it [organize materials] a little bit, and EndNote is trying to do it, but Zotero and EndNote don't do it so well. Papers does it far better because it not only has a kind of index structure to it, but it has a filing system, so that it actually files the PDFs away in your hard drive, in a kind of hierarchical folder structure.

In addition to citation managers, cloud-based services, which keep information on an Internet server, are also popular for storing and organizing research materials. One of the most popular is Dropbox, particularly among scholars in the humanities and social sciences. "My projects are now on Dropbox. I'm paying a monthly fee to get enough storage to do my work," says a

Survey respondents overwhelmingly indicated that they most frequently store information on the hard drive of their computer or computers.

Our interviewees indicated that they commonly store and organize research articles in an electronic format, often through the use of research or citation management software.

In addition to citation managers, cloud-based services, which keep information on an Internet server, are also popular for storing and organizing research materials.

professor of rhetoric. A professor of human development and family studies declares, "I use Dropbox for everything. It has saved my life, it has changed my life. I paid for a larger Dropbox storage so that I can put files to reside on Dropbox. But because I travel so much, it is the best thing ever." In like fashion, a professor of philosophy admits that he is "a huge fan of Dropbox."

Being popular storage solutions, cloud-based services are also often used for organizing materials. The practice of using the same tool for storing and organizing information can again be observed in the case of Dropbox. "I would organize it [research materials] into folders in Dropbox, and it would be folders by project or folders by coauthor, or folders by specific tasks," explains an associate professor of political science. Similarly, a professor of human development and family studies explains that her research materials are organized "just how I've set up my Dropbox. It's all Dropbox synchronized, and it's organized well enough for me to find stuff." Sometimes, the use of such commercial services is preferred to using university services: "I use Dropbox to house everything in the cloud," says an assistant professor of media studies. "I just don't trust the institution to not co-opt my work or look into it or keep it from me."

Other scholars, however, distrust commercial cloud-based services. For instance, an assistant professor of education explains: "I'm a critic of using many of the cloud services, in large part because of what I see are poor privacy issues." He specifies, "I put and share files in there when I feel I'm not concerned about the privacy."

In a similar manner, an associate professor of education explains that privacy issues and institutional review board (IRB) requirements oblige her team to use university instead of commercial services, which however causes problems in terms of both storing and organizing information:

Right now we're using ANGEL [the university course management software platform] because it's a secure site, so it will meet the IRB [requirements]. But it's just really awkward. It won't hold the big videos. There's no revision capacity within a file. Can't organize any of the videos there. That's a real problem. Unfortunately, with secure data, we cannot put things like that on Dropbox.

The extent to which privacy issues influence storing practices can best be observed in the case of scholars whose work requires such a high level of security that neither commercial nor institutional cloud-based services can be used for keeping research materials. For instance, an associate professor of education describes her storing practices in the following way:

It is hard drives and flash drives locked in my filing cabinet. It's under a password protection on the hard drive, and then it's under lock and key. Really, the biggest thing is security. If it wasn't completely secure, it couldn't be used. I don't actually know of anybody working with footage of elementary school students that doesn't use the lock and key method. Going through the IRB process, it's always been a caveat that it has to be under lock and key just because they're minors.

Data storing and management also stand as a big problem in the sciences. For instance, a professor of physics explains:



If I generate data or I need to make a copy of a large amount of data, right now the infrastructure that is available to me as a researcher at Penn State is not there. If I need to bring over a large amount of data, I basically have to buy the storage space and manage it myself. The ability to bring over and manage large data sets where I need to work intensively on them is just not there.

Scholars in the sciences see organization of stored materials as an additional problem, especially organization that would support data sharing and reuse. For instance, a professor of astrophysics explains:

It's one thing to say that I have it [stored data] and I can make it available to you, but is it organized? I could certainly give my data set to anybody, but without me there answering their question 24/7, it is useless to them. It is the documentation, organization, reliability, the metadata that make that information not simply useful, but usable in any way, shape or form.

Scholars in the sciences see organization of stored materials as an additional problem, especially organization that would support data sharing and reuse.

In addition to using cloud services, many scholars, especially in the humanities, store and organize research materials on flash drives, hard drives, or both, as well as in hard-copy form. For instance, a professor of French explains that she keeps research materials on computer hard drives and further manages them in the following way:

In addition to using cloud services, many scholars, especially in the humanities, store and organize research materials on flash drives, hard drives, or both, as well as in hard-copy form.

I have three computers and a USB key, which is my little, precious, precious tool. It's very complicated because when I work in one place, I have to make sure that the latest version of what I've done is transferred to each machine in due time. I do have all the versions of it on paper, as well. It's kind of an old-fashioned security thing.

Old-fashioned storing practices are often accompanied with old-fashioned organizing methods, which can be difficult when dealing with electronic materials. "I used to be better at organizing, and it became this volume of stuff. And I had a complete breakdown of my organization system," says an associate professor of education. The professor stresses:

Old-fashioned storing practices are often accompanied with old-fashioned organizing methods, which can be difficult when dealing with electronic materials.

In my past, not so long ago past, it was those paper copies that were really my source, my organization. It was open the file drawer and here's all the materials used for some



particular study. Things are now online. And that's really where I need to clean up and adjust my old organization system for the changing nature of that.

In a similar manner, a professor of French and gender studies says:

I actually have it [research materials] probably more organized at the level of the print materials than I do at the level of the digital. For one thing, I simply have changed computers over the years, so where I have not had an organized strategy has been to systematically move the materials from one computer to the other. That's really been a problem for me.

The use of "old-fashioned" research practices can also be observed with regard to citing information, which we elaborate in the following subsection.

Citing Information

Citation management is an important part of the scholarly workflow. Citation management programs can, with the click of a few keys, reformat citations in a myriad of citation styles and relieve the scholar of the arduous work of manually typing and reformatting citations. Yet, our study results indicate that the use of citation managers is rather low among Penn State faculty, in both the sciences and the humanities and social sciences. In our survey responses, slightly more than half (55 percent) of the faculty in the sciences reported using citation managers, while in the humanities and social sciences this percentage was lower, 30 percent. The tool most frequently mentioned in our survey results was EndNote, followed by Mendeley, Zotero, Papers, Sente, RefWorks, and BibTeX.

Our interview results somewhat differed from the survey results in terms of use and preferred citation management programs. Specifically, all of the interviewees in the sciences reported using citation managers, among which BibTeX and Papers were most common. Still, our interview findings confirmed major differences between scholars in the humanities and social sciences and those in the sciences with regard to the use of citation management programs.

Dissatisfaction with existing citation managers is a commonly articulated reason for not adopting citation management software. For instance, a professor of human development and family studies says: "I went to a class [on EndNote], I looked at Zotero; none of them work for me, so I just don't use it." Other respondents indicated that, although they did not use citation managers, they did think the use of such tools would be beneficial for their work.

Dissatisfaction with existing citation managers is a commonly articulated reason for not adopting citation management software.

or Zotero." A number of interviewees, mostly more senior humanities faculty, simply indicated not using any citation manager, without stating the reason or a wish to change such a practice.

"It's unbelievable that I don't use any of those [citation managers]. I still have my graduate assistant help me to draft my bibliographies. I can see that I would develop my bibliographies so much [more] easily and quicker," says an assistant professor of media studies. Similarly, a professor of English explains: "I do it by hand; that would be the chief reason for switching to Mendeley

Among our interviewees in the humanities and social sciences who do use citation managers, the most often used program is EndNote. Commonly, this is a tool that they have started using as graduate students and continued working with later in their academic career, although their general level of satisfaction with the tool is not high. For instance, an assistant professor of English as a second language says: "Am I satisfied with it [EndNote]? Not really. It's not very intuitive. Why I use it? I guess when I was in graduate school somebody recommended it to me and I bought it."

A few of our interviewees in the humanities and social sciences indicated using Zotero in addition to EndNote, highlighting some of Zotero's benefits. "I use Zotero now as my primary reference manager. Easy capture [is the major benefit of Zotero], and then easy retrieval," says a professor of rhetoric.

In contrast to scholars in the humanities and social sciences, all of our interviewees in the sciences indicated using citation manager software, most commonly BibTeX/JabRef or Papers. "We usually write all papers using LaTeX Bib-TeX Manager, and that's JabRef," says an assistant professor of chemistry. In the same way, an assistant professor of biology explains: "I write all of my articles in TeX, so BibTeX is a natural format, and there's a couple tools that relate to that."

A number of our respondents in the sciences also indicate combining citation managers according to their fit into a scholar's workflow. For example, a professor of astrophysics explains:

I have found that for my work there are two ways in which I use citations. One is when I'm writing; because I do all my writing in LaTeX, I want something that integrates well with BibTeX. So, I have a very large BibTeX library. [On the other hand] I'll get a question or I'll be responding in an e-mail to someone and I'll say, 'You really need to look at, or have you seen . . . ?' I can basically just drag the reference [from Papers] and it will come out in Harvard form [also known as parenthetical referencing] or something like that. And so, for that purpose, Papers or Mendeley is very, very useful. The difficulty, of course, is that it means that I have my library in two different places.

Storing electronic files at different places also came out as a challenge in our exploration of scholars' archiving practices, which we elaborate in the following subsection.

Archiving

In his book *History Hunting: A Guide for Fellow Adventurers*, the historian James Cortada says, "All researchers are archivists, whether they know it or not."⁹ This simple pronouncement highlights the essential need for information management training for faculty and graduate students. In our survey, users were asked about their knowledge of, and general practices related to, saving important information. A total of 89 percent of faculty respondents reported that they make backup or additional copies of important materials. The most frequent location for storing backup copies was the user's external hard drive (67 percent) or a flash drive (49 percent); followed by a personal computer hard drive (49 percent); printouts (36 percent); a cloud server, such as Dropbox or Google Drive (33 percent); and a personal or departmental server (18 percent). In general, storing practices and backup practices were very similar, and use of institutional storage to preserve significant items was not prevalent among respondents. Respondents were

also asked whether they regularly destroyed or removed files that were no longer of use; a majority (56 percent) indicated that they do not do so.

All of our interviewees in the sciences, and most of those in the humanities and social sciences, identified archiving as one of their regular scholarly practices.

In terms of interview results, our study showed both similarities and differences in scholars' archiving practices across academic disciplines. All of our interviewees in the sciences, and most of those in the humanities and social sciences, identified archiving as one of their regular scholarly practices. Our interviewees indicated flash drives and cloud-based services as their most often-used archiving tools.

For instance, an assistant professor of chemistry explains that he uses Dropbox for archiving his daily research materials, and an external hard drive for storing materials to which he does not need immediate access. He explains:

This [writings, notes] is automatically archived into Dropbox. Then all the data, correspondences, papers, and grants that I write, not the papers I've collected to read, goes through a program called SugarSync. It just creates automatic version backups to a cloud service that's accessible from anywhere. Papers [downloaded articles] happens to be a huge file, it's maybe 10 or 12 gigabytes, and that I back up to an external hard drive using TimeMachine on the Mac because I rarely need immediate access to that in any given time.

Some of our interviewees, however, consider physical devices as more reliable archiving solutions than cloud services. For instance, an assistant professor of English as a second language specifies: "Since I'm saving things on Dropbox, mostly everything is there and then every few months I try to back it up on an external drive."

A number of faculty indicated that sharing information on the Web is a way of archiving materials. "My blogs are a kind of place for archiving," explains a professor of philosophy. "I have a neat portfolio that every time I do something, I write a blog post about it, and if I [am] publishing something, so that's really a good archive," our respondent specifies.

A number of faculty indicated that sharing information on the Web is a way of archiving materials.

For scholars in the humanities and social sciences, dealing with obsolete technological formats has been highlighted as an important challenge of digital archiving.

For instance, an assistant professor of education explains: "I've got tape backup stuff from 1990s, but I have no machine to read it. I've got five and a quarter inch floppy drives, I've got three and a half inch floppies. This stuff is in my garage. I'm not deleting it, but it's getting trapped in that form." Similarly, a professor of English specifies: "I've lost a bunch of files, which didn't make the transition. It is a major problem. That happens, but it's something that I don't think should have happened."

A number of interviewees in the humanities and social sciences indicated not archiving their research materials at all, for different reasons. Sometimes, the reason is a

lack of habit, skills, or both: "Honestly, it's really up to graduate students. I would not personally be able to keep up with doing it at this point," says a professor of human development and family studies. Similarly, an assistant professor of French and linguistics admits: "Big confession, no backups. It's insane. I know it is [dangerous]. I don't have any external storage devices because I don't know how to use them. I'm just absolutely ignorant about those [cloud-based services]." Our respondent also explains that she often prints out electronic texts and stores them as printed files.

In other instances, the lack of archiving is related to privacy issues, as also mentioned when we discussed scholars' storing practices. "Most of my data is archived in my basement," says an associate professor of education. The professor continues:

I have them all labeled, so after something's been published five years, I can destroy it. We have videos of children who are identified in our studies and easy to trace back, identified with emotional behavior disorders. The chance that that student video would ever get out and that child would ever have the whole world know, it's very frightening to think of that.

For some of our respondents, digital archiving is preconditioned with the need to digitize their research materials. For instance, an associate professor of political science explains:

Back in the 1980's, I did a lot of fieldwork and collected a lot of government documents, interviewed a lot of people, had surveys of people, and all of that is on paper. In terms of archiving, it would be nice if it were possible to inexpensively and quickly scan so that there would be a digital copy.

Finally, an important difference between humanities and social sciences scholars on the one hand and those in the sciences on the other could be observed in the approach toward archiving their research traces. Namely, while the former primarily focus on archiving final research outputs (publications), the latter are concerned about archiving their research trail and data too.

For instance, a professor of physics stresses that it would be highly beneficial to archive and make available and searchable data sets, research notes, lab books and similar materials, but that the tools and resources for doing that do not exist. Our respondent explains:

I've had requests from people, "Do you have that code available?" I have printouts of that code, but God help us all if our lives depended on my being able to get that code to compile and run again. So, archiving is more than just making a copy available somewhere. Making something available but not in a way that it's at all useful, it's a waste of resources. I certainly don't know how to do it well, and I don't know of anybody else in my discipline that does it well.

In a similar way, a professor of biology posits that it would be useful "if there was a trail of how people came to the conclusions that they did," adding, however, that "to do that, you'd have to have huge amounts of storage space."

Among our interviewees in the humanities and social sciences, a professor of rhetoric stands out in his focus on preserving research traces. "Core is the research trail part,



how do we preserve the research trail, so that we can see where we went wrong, or see discovery that we didn't notice we were making at the time, and go back and recover it," our respondent stresses. "If we obliterated all that, we would have lost something valuable to every discipline—disciplinary knowledge," he concludes.

After presenting our results related to finding, storing, organizing, citing, and archiving information, in the next and concluding part of this section we turn to the results related to the study participants' attitudes toward the role of the library in their work.

The Role of the Library

With regard to the role of the library in scholars' work, our survey questions focused on identifying the areas of instruction that the library could provide, while our interview questions examined the broader role this institution has in the respondent's research practice.

While participating faculty agreed that it was easy to find research articles (88 percent), store them (85 percent), cite (81 percent), save or archive (79 percent), and share them (69 percent), half or less than the half of participating faculty felt the library should have a role in instructional support relative to the research workflow. For example, 53 percent indicated that the library might offer training programs in citation management software. While the Penn State Libraries offers on-site training for EndNote, Zotero, and Mendeley, this interest in citation management training may speak to a need for greater promotion of these classes across campus, for providing instruction on other citation management programs, or both. A total of 44 percent of faculty also felt that the library could offer training in how to use research databases, including Google Scholar. Such training is also happening at Penn State, but perhaps not on as granular a level (discipline-focused) as it might need to be to catch faculty's attention or to be useful for their work. Faculty were also interested in personal archiving assistance (45 percent), speaking to a need that has yet to be addressed broadly in many academic libraries. Less popular but still noted by some faculty as areas of training needed were applications for organizing research, such as wikis, RSS [Rich Site Summary, often called "really simple syndication"] feeds, which update syndicated data automatically, or blogs (39 percent). Respondents also expressed interest in learning management of research assets / research data (38 percent) and in training on how to use Penn State's institutional repository, ScholarSphere (31 percent).

Multiple faculty declared the opinion that learning how to use research-related technologies was the responsibility of the scholar, not that of the library. Open-ended answers that the surveyed faculty provided to the question "What training programs should University Libraries offer?" included responses such as: "None; it's the scholar's

responsibility to learn how to deal with these issues"; "This stuff is my job as a researcher"; "It varies with discipline"; "None—training would not be specific enough for the discipline"; and "None, most of these are self-explanatory and the rest are too case-specific."

The majority of our respondents specify that electronic search and access have replaced physical visits to the library.

In the interview results, although our respondents unanimously agree that their interacting with the library and library collections has significantly changed in the digital age, they also agree that the role of the library has remained the same. In terms of interacting with the library, the majority of our respondents specify that electronic search and access have replaced physical visits to the library. For instance, a professor of biology explains that he “used to come to the library at least once a week to check the journals, but with the electronic journals [he doesn’t] do that anymore,” while an assistant professor of kinesiology says: “I’ve probably not been to the library very much in the two and a half years that I’ve been here [at Penn State], but I do use a lot of the online resources; that’s primarily what I use the library for.”

In terms of library functions, our respondents see the library as a conduit of information, specifying that the role of libraries is to provide scholars with an easy access to information. “What I need is quick and efficient access to existing materials, be they print or digital,” says an assistant professor of biology, stressing: “The faster and easier it is for me to get those resources,

Our respondents see the library as a conduit of information, specifying that the role of libraries is to provide scholars with an easy access to information.

the more productive I can be.” In a similar manner, an assistant professor of mathematics specifies that “The researcher wants you [the library] to help provide him with access to other researchers’ work so he can do his own research better,” while an assistant professor of engineering design highlights that the role of the library is “helping us find resources, helping the content become accessible.” This finding reaffirms results of the Ithaka S + R US Faculty Survey 2012, in which faculty saw the library’s most important role as buyer, repository, and information gateway.¹⁰

In general, our respondents’ outlook on the role of libraries in the digital age is well summarized in the following quotation from a professor of astrophysics:

The only thing that has really changed about the role of the library is as a physical repository for local copies of things. But all of the functions that I used the library for before and that I relied on the library for, none of those functions have gone away. They’re just being delivered in a different fashion now, and one that doesn’t require me to physically go to use them . . . The physical aspect of the library is not necessarily important for contemporary information, but the cataloging of the information, the metadata, the aggregation, the ability to search, all of those functions which libraries provide are even more essential.

An emerging role of the library that our respondents see or would like to see is related to data management and digital archiving.

In addition to these traditional functions of the library, an emerging role of the library that our respondents see or would like to see is related to data management and digital archiving. For example, an associate director of strategic interdisciplinary research explains:



I think of the libraries in two separate roles. My most frequent use of the library is journal access. Next to that, work on data management and their role in archiving information. I thought of the library's archiving as old stuff, maps and paper things from way back. I suddenly became to think of them as electronic digital archiving.

In a similar manner, an assistant professor of chemistry says, "One [role of the library] is making access to journals and search tools, which is currently what it is, [but] it could be helping with data management plans," while a professor of biology comments: "I like the idea about [library] archiving and storing data. If you don't have to store books in the library, but instead, you have a computer that's storing information, I think that's sort of the same thing."

With these findings about the role of the library, we wrap up our results section, and we now turn to discussing those findings and to offering our recommendations and conclusions.

Discussion and Conclusion

The results of our study indicate that digital technologies have different roles and levels of integration at various phases of scholarly workflow. For instance, academics actively use digital tools for finding, storing, and archiving research materials, but infrequently

Academics actively use digital tools for finding, storing, and archiving research materials, but infrequently use them for creating bibliographic citations.

use them for creating bibliographic citations. Although the use of citation management programs is somewhat higher in the sciences than in the humanities and social sciences (55 percent versus 30 percent), the overall level of digital technology use in this research activity is lower than in other phases of the research workflow. Similarly, while our respondents perceive as useful innovations digital tools for find-

ing, storing, and organizing information in their research, employing digital tools for creating bibliographies is often perceived as a hurdle, particularly in the humanities and social sciences. This finding might be additionally surprising given that digital citation management has been part of university services and library training programs for a number of years, and the majority of our respondents indicated early (that is, graduate level) adoption of these tools. This finding thus implies that although institutional support and training programs are vital for the uptake of digital tools, such programs are not necessarily sufficient for effective integration of those tools into scholarly practice.

Our results further illustrate various ways in which integration of digital tools in one phase of the research process influences other segments of the workflow. For example, scholars' full reorientation on electronic search and access produces an abundance of collected materials, requiring adjustments in researchers' storing, organizing, and archiving practices. As some of our respondents observed, integration of digital tools into their search activities resulted in a complete breakdown of their systems for organizing information, which were developed for print-based materials. Therefore, while implementation of digital tools into one phase of the workflow might be rewarding, it might

also become a challenge in other phases of the work. This finding implies that digital research tools should be designed to support a continuous research workflow instead of separate and disconnected activities.

Our findings also suggest that in the workflow of a digital scholar, technical rather than traditional methodological expertise shapes interconnectedness among phases of the workflow. For instance, in our study, greater level of workflow interconnectedness was observed among scholars in the sciences, who tend to be more technologically savvy than academics in the humanities and social sciences. For our respondents in the sciences, finding, storing, organizing, and sharing information often seemed like one unified activity, commonly performed through the use of one digital tool (for example, Papers). Furthermore, our respondents in the sciences tend to require better integration of different phases of the research process, such as insisting that a storage service include annotating and sharing functions to be useful.

This, as well as our other study findings, indicates a significant scope of disciplinary differences with regard to the use of digital technologies in scholarly work. Broadly conceived, these disciplinary differences can be conceptualized as either inherent or acquired. As an example of inherent disciplinary differences, we could understand data privacy requirements, which widely differ across disciplines and, as our findings show, significantly determine the type and level of digital technology use. Similarly, vast data sets in the sciences versus small or relatively small data sets in the humanities are instances of inherent disciplinary differences that shape digital technology use. Acquired differences, on the other hand, can be observed in a set of habits and assumptions rooted in a particular community of practice. For instance, persistence of disciplinary habits can be seen in humanities and social sciences scholars' tendency to keep using the same citation management programs despite being dissatisfied with them. As our findings show, although a citation management program such as Papers meets the desired functionalities of academics in the humanities and social sciences, they keep using EndNote or Zotero, either because they are not aware of other citation management programs or because institutional support endorses a particular tool. As Collins and her coauthors point out, "Disciplinary traditions and learned behaviours determine the types of digital tools and services that are developed for or by researchers."¹¹

This finding also implies another disciplinary difference, which is that humanities and social sciences scholars might be more reliant on institutional assistance and instruction, while researchers in the sciences seem to be more prone to individual discovery of, and experimentation with, digital tools. This observation should, however, be taken in relative terms, as higher levels of experimentation with digital tools could also be found among some of our humanities and social sciences respondents. For instance, the use of enhanced digital publications to stimulate dialog among humanities scholars represents one such example. Still, some perhaps surprisingly insufficient levels of technical competency could be observed among a number of our humanities and social sciences respondents. For example, unawareness of the basic cloud storage services or reliance on printing out materials to archive them could be seen as evidence of a lack of technological savvy. In general terms, the humanities and social sciences scholars expressed more difficulties in adjusting their research practice to the requirements and possibilities of digital scholarship, such as developing organizational and citation management

methods suited to electronic materials, migrating their research materials from obsolete to contemporary formats, and digitizing research materials.

As previously mentioned, technical architecture of digital research tools needs to support specific disciplinary needs in ways that address both inherent and acquired

While storage capacity stands out as a challenge in the sciences, high security and privacy requirements characterize needs in the social sciences.

disciplinary differences. Data storage and management, for instance, has been identified as a serious problem across disciplines, but with distinctive disciplinary needs. While storage capacity stands out as a challenge in the sciences, high security and privacy requirements characterize needs in the social sciences. This challenge is highlighted by the fact that academics sometimes depend on institutional repositories for data storage and management, which tend to lack storage capacity, revision and annotation functionalities, and so on. Due to

such shortcomings, as our results show, scholars commonly use commercial storage services when not administratively required to stay within an institutional framework.

Advantages of commercial over academic tools and services can also be identified in other examined segments of digital workflow. As we saw in the previous section of this paper, many academics prefer Google and Google Scholar for information search, in the same way as they elect to use Dropbox for storing and Papers or EndNote for reference management. This finding is consistent with other recent studies, implying that academic tools and services need to address user needs better to become more competitive with commercial services.¹²

Survey respondents were asked about the ease and challenges of their current research workflow, in hopes that the findings would underscore areas of need with regard to training and outreach. While a majority of users felt that it was easy for them to find, store, cite, and share information, many respondents reported difficulty with managing files. For example, 44 percent of respondents indicated that they had lost a file or other important information stored electronically. Nearly a third of respondents (28 percent) also said that they had lost files due to outdated or obsolete formats. These responses point to an area of significant focus for librarians and technologists—unifying the research workflow for users. Loss of information is significant, and there is a need to build into the research workflow easy strategies for users to self-archive their work in storage services that are inherent to the individual or the institution. These provisions could include institutional repositories or institutional subscriptions to cloud storage services. The high rate of users in this study losing information (and saving information on computer hard drives, flash drives, or both) mandates locally based action to help save critical information for both the individual and the institution.

Similarly, discovery must also be better handled for the end user. While users state ease of use with regard to finding information, there is still information loss because storing and organizing information is a process separate from retrieval. Libraries' current dependence on Web browser-based search is not sustainable in the long term. Search and discovery must migrate into the users' environment and their natural workflow. Current bibliographic management software such as Papers and ReadCube bring this idea to

life. Search and retrieval is fully integrated into an interface that also allows annotation, organization, and archiving of PDFs and other research materials.

The next proposed phase of this study will focus on both of these areas: self-archiving and discovery, using Zotero as a test platform. This phase will explore effective methods to connect the institutional repository within Zotero, as well as to expose references and metadata within uploaded PDFs, with the goal of further unifying several phases of the research workflow, and better integrating finding and archiving into the scholar's online path.

Finally, our results show that in the digital age the role of academic libraries does not diminish but actually becomes stronger, albeit in a modified form. Our findings suggest that, instead of trying to reinvent themselves, academic libraries should continue doing what they have been doing, providing scholars with expertise in the areas of information management, storage, and access. A transition to digital information management, storage, and access gives college and university libraries an important role in digital scholarship; as one of our respondents says: "The libraries are becoming even more central than they ever have been, and part of that involves making sure that the collections are accessible from distance and are able to be interacted with in a dynamic way . . . to integrate itself into the research circle of every faculty member."

Instead of trying to reinvent themselves, academic libraries should continue doing what they have been doing, providing scholars with expertise in the areas of information management, storage, and access.

Smiljana Antonijević is an assistant professor of culture and technology in the Department of Culture and Identity at Roskilde University in Copenhagen, Denmark; she may be reached by e-mail at: smiljana@smiljana.org.

Ellysa Stern Cahoy is education and behavioral sciences librarian and assistant director of the Pennsylvania Center for the Book at the Penn State University Libraries in University Park; she may be reached by e-mail at: ellysa@psu.edu.

Notes

1. Jo Bell Whitlatch, "Library Use Patterns Among Full- and Part-Time Faculty and Students," *College & Research Libraries* 44, 2 (March 1983): 141–52.
2. Susan V. Wawrzaszek and David G. Wedaman, "The Academic Library in a 2.0 World," ECAR [EDUCAUSE Center for Analysis and Research] *Research Bulletin* 19 (2008), <http://net.educause.edu/ir/library/pdf/erb0819.pdf>; David Nicholas, Ian Rowlands, Peter Williams, David Brown, and David Clark, "E-Journals: Their Use, Value and Impact—Final Report," Research Information Network, London (2011), <http://www.rin.ac.uk/system/files/attachments/E-journals-report.pdf>.
3. David Seaman, "Discovering the Information Needs of Humanists When Planning an Institutional Repository," *D-Lib Magazine* 17, 3–4 (March–April 2011): 1–15, <http://www.dlib.org/dlib/march11/seaman/03seaman.html>.

4. Catherine Marshall, Sara Bly, and Francoise Brun-Cottan, "The Long Term Fate of Our Personal Digital Belongings: Toward a Service Model for Personal Archives," *Proceedings of IS&T* [Society for Imaging Science and Technology] *Archiving* (2006): 25–30; Dorothea Salo, "Innkeeper at the Roach Motel," *Library Trends* 57, 2 (Fall 2008): 98–123 http://muse.jhu.edu/content/crossref/journals/library_trends/v057/57.2.salo.html; Roger C. Schonfeld and Ross Housewright, *Ithaka S + R US Faculty Survey 2012* (April 8, 2013), <http://www.sr.ithaka.org/research-publications/us-faculty-survey-2012>.
5. Susan Siegfried, Marcia J. Bates, and Deborah N. Wilde, "A Profile of End-User Searching Behavior by Humanities Scholars: The Getty Online Searching Project Report No. 2," *Journal of the American Society for Information Science* 44, 5 (June 1993): 273–91; Lokman I. Meho and Helen R. Tibbo, "Modeling the Information-Seeking Behavior of Social Scientists: Ellis's Study Revisited," *Journal of the American Society for Information Science and Technology* 54, 6 (April 2003): 570–87; Allen Foster, "A Nonlinear Model of Information-Seeking Behavior," *Journal of the American Society for Information Science and Technology* 55, 3 (February 2004): 228–37; Bradley M. Hemminger, Dihui Lu, K. T. L. Vaughan, and Stephanie J. Adams, "Information Seeking Behavior of Academic Scientists," *Journal of the American Society for Information Science and Technology* 58, 14 (December 2007): 2205–25; Carole L. Palmer, Lauren C. Tefteau, and Carrie M. Pirmann, "Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development," report commissioned by OCLC [Online Computer Library Center] Research (2009), oclc.org/programs/publications/reports/2009-02.pdf; Ellen Collins, Monica E. Bulger, and Eric T. Meyer, "Discipline Matters: Technology Use in the Humanities," *Arts and Humanities in Higher Education* 11, 1–2 (February 2012): 76–92.
6. A number of interviewees opted for being interviewed in the Penn State main library complex.
7. Nicholas, Rowlands, Williams, Brown, and Clark, "E-Journals: Their Use, Value and Impact—Final Report."
8. Simone Kortekaas, "Thinking the Unthinkable: A Library Without a Catalogue—Reconsidering the Future of Discovery Tools for Utrecht University Library," Association of European Research Libraries (2012), <http://www.libereurope.eu/blog/thinking-the-unthinkable-a-library-without-a-catalogue-reconsidering-the-future-of-discovery-tools>.
9. James W. Cortada, *History Hunting: A Guide for Fellow Adventurers* (Armonk, NY: M. E. Sharpe, 2012), 165.
10. Schonfeld and Housewright, *Ithaka S + R US Faculty Survey 2012*.
11. Collins, Bulger, and Meyer, "Discipline Matters," 82.
12. Kortekaas, "Thinking the Unthinkable."