Researcher as Bricoleur: Contextualizing humanists’ digital workflows

Smiljana Antonijevic <smiljana_at_smiljana_dot_org>, The Pennsylvania State University
Ellysa Stern Cahoy <ellysa_at_psu_dot_edu>, The Pennsylvania State University

Abstract

This paper presents findings from a set of contextual inquiry sessions completed in 2016 as part of the four year Digital Scholarly Workflow project (2012-2016) conducted at the Pennsylvania State University and funded by the Andrew W. Mellon Foundation. Studying humanists’ workflows through contextual inquiry enabled us to gain valuable empirical insights into digital scholarship from the perspective of individual scholars. This small-scale observational study (N=8) supplemented findings from our larger scale (N=196) and medium scale (N=23) studies previously completed in the project, allowing us to develop deeper qualitative understanding of how humanists’ research practices change in the encounter with digital technologies. Through close contextual analysis of information-based research activities we illustrate how humanists build rich personal collections of digital sources, which become important information resources in their work. We also examine how adoption of digital tools transforms the established methods of managing primary and secondary sources, bringing hastier digital workflows to traditionally more methodical work. Finally, we show that efficiency cannot be qualified as a neutral or inherent characteristic of digital tools, independent of scholars’ practices, since scholars’ interaction with digital artifacts determines what constitutes efficiency of a tool in the context of a scholar’s workflow. We conclude that in humanists’ workflows digital and analog tools and resources do not necessarily replace one another or vie for the dominant position. They rather supplement each other in an interconnected, hybrid workflow that comes to our respondents as an organic manifestation of their work, and as an illustration of the bricolage character of humanists’ digital workflows.

Introduction

Discussions of digital scholarship in the humanities have often been based on assumptions about what humanists could do with digital technologies, and how these technologies could influence their research, rather than on contextualized analytical insights about what humanists are actually doing and how their research practices change in the encounter with technology.

This paper presents the results of the Digital Scholarly Workflow project, which implemented an ethnographic approach to facilitate such contextualized understanding of humanists’ research workflows, giving us valuable empirical insights into digital scholarship from the perspective of individual scholars — how they search for, find, collect, analyze, and when necessary digitize archival and other scholarly materials. Our research approach builds on [Svensson 2010], who called for the examination of digital scholarship from the perspective of individual scholars’ daily practices, and [Kirschenbaum 2014] who argued for the analysis of digital humanities in action. Schonfeld (2015) similarly pointed out that “to understand researcher practices, user experience specialists … should examine the researchers’ actual practices,” for which “[taking], a more holistic, ethnographic perspective is vital” [Schonfeld 2015, 13].

The Digital Scholarly Workflow project was conducted at the Pennsylvania State University from 2012 to 2016 and funded by the Andrew W. Mellon Foundation. The first part of the project (2012-2014) analyzed how scholars across disciplines engage with digital technologies at different phases of the research lifecycle, pointing at gaps in existing tools, services, and infrastructures (see: [Antonijević 2015]; [Antonijević and Cahoy 2014]). While acknowledging the importance of domain-based and subject-specific research practices across the humanities (see: [Wineburg 2001]), our
study sought to understand some of the common features of humanists’ digital workflows as compared to those in the sciences, and to identify the distinctive features that software architecture should have in order to support humanists’ research.

The first phase of our project thus identified preferences and areas of strain in the research workflows — specifically around discovery, organization, and storage of research materials — and revealed distinct disciplinary differences. For instance, humanities scholars preferred using local discovery tools, such as the Penn State library catalog, more often than Sciences faculty. Humanists also noted a lack of digital technology use in data collection and analysis phases of their research process. Further, humanities scholars had the highest percentage of lost and inaccessible research files, and were more likely than other disciplines to curate their research file collections. Finally, phase I showed that integration of digital tools into humanists’ search activities often resulted in a breakdown of their systems for organizing information, previously developed for print based materials [Antonijević and Cahoy 2014].

Building on those findings, phase II of our study (2014-2016) focused on observing humanists’ research and use of digital research tools in situ, through a series of contextual inquiry sessions. Contextual inquiry showed that digital technologies have transformed scholarly work in the humanities not only as reflected in the discourse of digital humanities, but through daily practices of working scholars and through some perhaps surprising integrations with long standing analog work routines. In this paper we demonstrate, for instance, how impromptu on-site digitization of archival materials with the use of digital cameras change scholars’ research in both rewarding and challenging ways. Through such practices scholars build rich personal collections of digitized primary sources, which become one of the central information resources in their work. Our findings show, however, that digital tools and resources do not necessarily replace analog or compete for the dominant position, but form an interconnected stream, illustrative of the bricolage character of humanists’ digital workflows.

Our view of humanists as bricoleurs builds on conceptual understanding of bricolage developed over the past fifty years. While the notion of bricolage as a process and bricoleur as an agent derives from Lévi-Strauss’s (1962) analysis of mythical thinking, it assumed broader epistemological meaning already in this original account. Soon after, and thanks to its broad generative capacity, bricolage attained the status of a universal concept applied across knowledge domains as diverse as evolutionary biology [Jacob 1977], interpersonal relationships [Conville 1997], nursing [Gobbi 2004], entrepreneurship [Baker and Nelson 2005], digital culture [Deuze 2006], oncology [Broom 2009], migration studies [Craciun 2013], political science [Carstensen 2011] and so on. The concept of bricolage also became used in the context of scholarly work, specifically related to interdisciplinarity and multiperspectivism in the adoption of research methods, data sources, and theories (see: [Denzin and Lincoln 2000]; [Kincheloe 2001]).

Our analysis of bricolage in humanists’ scholarly work does not follow this trajectory of explicit methodological and epistemological examination, nor does it consider Derrida’s (1997/1967) approach to bricolage as discourse. Rather, we seek to return to Lévi-Strauss’s inquiry into bricolage as situated practice and its specific ideology of unscripted approach to technology expressed in the comparison between engineers and bricoleurs discussed in more detail in the conclusion of this essay (see also: [Johnson 2012]). Along these lines we argue that humanists observed in our study engage with digital tools in a highly personal, unscripted way, which corresponds to their style of thinking and working, to their intimate epistemology. Our respondents do not experience digital workflow as a technical activity, but as a subjective process of knowledge production in which they freely combine diverse tools and resources suited to their sensibilities and research needs. Such use of digital technologies also emerges from their understanding of efficiency as contextualized in scholars’ interaction with digital artifacts, not as a presumably inherent characteristic of digital tools independent of scholars’ practices. Certain methodological and epistemological inferences that we offer in the conclusion of this essay thus seek to inform DH theory via DH praxis, in line with the orientation of our study.

Such focus on DH praxis was essential in phase II of the Digital Scholarly Workflow project, in which we partnered with the Zotero software (see: https://www.zotero.org) development team from George Mason University to enhance Zotero based on our research findings. Two software optimizations were developed and launched, centered on discovery, organization, and archiving in the research workflow. A Really Simple Syndication (RSS) feed reader was introduced in Zotero that allowed researchers to receive notifications from their favorite journals and other publishers as new
publications appear. A Zotero “My Publications” workflow was also created, offering researchers the capability to self-archive their work. These items could then be automatically deposited into a Hydra-based institutional repository via a new tool, Arkivo, developed for the project (see: https://github.com/zotero/arkivo-sufia). The second phase of the Digital Scholarly Workflow project also included usability testing of these optimizations, indicating positive findings towards centering specific research activities within the Zotero interface for greater workflow productivity and cohesiveness (see: [Cahoy 2018]).

We start this essay with a summary of the Digital Scholarly Workflow project methodology, and then turn to presenting the results gathered through contextual inquiries. Specifically, we focus on results related to the research activities of finding, organizing, and storing digital research data and materials, which were the focal point of phase II of our project. The initial project phase, which analyzed scholars’ information behavior at all stages of the research lifecycle, showed that the important links and mutual influences among these research activities are not always fully recognized, so, in phase II, we sought to contribute to filling in those analytical gaps. The paper concludes with the vision of scholars as bricoleurs, and with suggestions on how to approach digital scholarship in the humanities based on ethnographic understanding of digital workflows in context.

**Project methodology**

Phase I (2012-2014) of the Digital Scholarly Workflow project explored the digital research workflows of the Penn State faculty across the sciences, social sciences, and humanities, and it included a web survey (N=196) and in-person ethnographic interviews (N=23). The quantitative orientation of the web survey enabled us to compare and contrast what scholars are doing in their digital workflows, while the qualitative character of the in-depth interviews facilitated understanding of why scholars carry out their workflows in particular ways.

Building on those results related to the “what” and “why” of digital workflows, phase II of the project (2014-2016) focused on humanities scholars and on understanding a twofold how – how humanists’ carry out their workflows in practice, and how software solutions can enhance or impede those practices. As mentioned in the previous section, we observed humanists’ use of digital research tools in situ through a series of contextual inquiry sessions, with the goal of answering the first how, and we then partnered with the Zotero team to address the second how by enhancing the Zotero software. Our Zotero enhancement activities are addressed in detail in separate articles (see: [Antonijević and Cahoy 2016]; [Cahoy 2018]), while this paper focuses on the results of the contextual inquiry sessions completed in the Fall of 2016.

Contextual inquiry is a qualitative method that combines traditional ethnographic techniques into a potent heuristic for gathering data about contextualized and naturally-occurring activities (see: [Holtzblatt and Jones 1993]; [Bednar and Welch 2014]). Context refers to a set of interconnected conditions in which analyzed activities take place, and which motivate and shape those activities. In this study, contextual inquiry concentrated on the digital workflows of eight humanities scholars in the course of their daily research. This small-scale observational study thus supplemented findings from our preceding larger and medium-scale studies completed in phase I, allowing us to develop deeper qualitative understanding of how humanists’ research practices change in the encounter with digital technologies. Scholars consulted in the contextual inquiries included four male and four female participants, most of whom (6) were tenured professors. More than half of the contextual inquiry sessions (5) were conducted at scholars’ offices and on average lasted 90 minutes.[2]

Two researchers attended each of the sessions and used the same observational guide, which provided for the consistency and easier standardization of research notes and findings. These researchers quietly sat aside observing study participants as they carried out their work, interrupting only if a clarification was needed, and taking detailed notes of the observed workflows. We did not assign any tasks to the participants, but asked them instead to engage in the information-based research activities they planned for the day, and to guide us through those activities by thinking aloud about why they are doing what they are doing. The respondents’ verbal accounts were audio-recorded and later professionally transcribed. Both the session transcripts and the researchers’ notes were coded for activities, patterns, themes, and respondents’ experiences in interacting with digital tools. All information that could have been identified with individual study participants remained confidential. The quotations used in this text are thus presented through
pseudonyms, and some of the quotations have been slightly edited for privacy, clarity, and length. In the following, we first discuss how our respondents search for research data and materials, and we then turn to reviewing their activities related to storing and organizing those materials.

Searching for research data and materials

When searching for research data and materials, humanities scholars consult a web of resources, both electronic and paper-based. This study focuses on electronic resources for two reasons. First, the study explored digital workflows, that is, the integration of digital tools and resources into humanists’ practice. Second, contextual inquiry allowed direct observation of scholars’ use of online resources, while the information on their use of libraries, archives, and other physical collections could have been inferred only indirectly, through the respondents’ accounts.

Contextual inquiry showed that humanists predominantly use online resources to search for secondary sources, such as journal articles and books. In discovery search, they frequently use commercial search engines like Google and Google Scholar, while library portals and academic databases mostly serve for known item search and access. Contextual inquiry thus confirmed the findings of phase I of our project and other relevant studies (see: [Antonijević 2015]; [Antonijević and Cahoy 2014]; [Nicholas et al. 2011]; [Asher, Duke, and Wilson 2013]). Humanists also explore more specialized online resources, such as movie databases and newspaper online archives, as well as their personal collections of digital and/or digitized materials. Among our respondents, the search for primary sources largely takes place offline, in archival and other physical collections, and to “digitize material” typically means to take a digital photo of an archival document and to store it as a JPEG file for further use. Such on-site digitized primary materials grow over time into humanists’ rich personal collections, becoming a vital information resource these scholars regularly access in their research.

For instance, in one of the contextual inquiry sessions, Audrey, a professor of history, searched for literature on an event that took place in 1916, and for which she had only partial information. Audrey’s search starts with her personal collection of notes written in Word and stored on the internal hard drive. She uses a Word search function that queries the folder for a supposed event name, but this search yields no result. Audrey then switches to her browser and the online search. She logs on to the Penn State library and enters a search phrase composed of three descriptors into the discovery search interface, LionSearch. This attempt does not yield any results either. “Okay, no problem, I’m going to go to some of my favorite databases,” Audrey says optimistically, and, using the same search phrase, she continues her search in the Historical Abstracts database. “All right, I need another field. It happened in Rome,” she comments still optimistically, and expands her search with one more field, which reads “Rome.” Still nothing. “Seriously?!,” Audrey exclaims with annoyance. “All right, let me just do ‘war council,’ something more specific,” she says with reasserted optimism, and changes her search phrase accordingly. Failure again. “Really?!,” Audrey laments in shock. “I would have thought it was more important.” Audrey then reaches to her bookshelf and grabs a book. She reads through a few pages, trying to find any additional information that could help her search. Nothing. But Audrey is not ready to give up yet.

She returns to her library search and adds “November” as one more search field, trying to make her query as precise as possible. No results. Still, Audrey does not give up, and, instead of adding one more search term, she decides to change her search phrase. She creates a new search phrase, again composed of three descriptors as the possible event name. “Nope. All right, strange,” Audrey says quietly, confident that any further search would be pointless. “You would think someone must have written an article about this. It was the time that the different allies got together and hammered out a strategy …,” she continues murmuring, but discontinues her library search. Instead, Audrey decides to try her luck with Google Search. She enters the search phrase and the Wikipedia entry pops up right away. “See, that’s the thing,” Audrey comments. “One would love to use more scholarly resources, but I just typed [the search phrase] and it’s up there [on Wikipedia]! Sadly, Historical Abstracts was not of too much use; the most useful one was still Wikipedia,” this historian concludes.

Dominance of non-academic search services was also observable in the case of Theodora, an associate professor of architecture, whose search for publications on housing in the Soviet Union started directly with Google Scholar, omitting
the library portal. Theodora starts by opening the Google Scholar page in her web browser and entering search keywords, “Soviet Housing.” Before looking at the search results, Theodora reaches to both the Zotero and Papers citation managers and creates a new collection in each of them, which she will use to store the materials found online. Theodora creates the collections manually, since she does not “trust the program to use its smart features.”[3]

Here we see that activities of finding and storing research materials can be so intertwined in scholars’ practices that setting up a storage folder goes in parallel, or even precedes search. Through her workflow and narration in the contextual inquiry, we learn that Theodora uses Zotero for saving bibliographic information about books, and Papers, for which she has paid subscription, for saving downloaded articles. Theodora relates that she particularly appreciates the Papers activity tab, which she calls “the magic finding feature,” which recommends further readings based on the articles she stored, and also enables her to discover tweets and other posts about those articles. Yet, despite highlighting the “magic finding feature,” Theodora does not use Papers’ built-in search tool, but carries out her searches through Google. This also influences the way in which she imports materials to Papers. Instead of using the built-in import tool, Theodora downloads files from Google and saves them on her desktop, after which she moves them to Papers. This rather untrained use of Papers is a confirmation of findings from the first phase of our study, which revealed that only a modest percentage of humanists use citation managers — one third (33%) of those surveyed — and, when they do use them, they employ them on a rather basic level (see: [Antonijević 2015]; [Antonijević and Cahoy 2014]).

After setting up new collections in both citation managers, Theodora returns to the web browser and starts reading through the results that Google Scholar yielded. “This one – it’s an edited volume, so I’ll look at that later. It is from 2005 and it’s talking about the Soviet Union, which is a little bit weird, but it looks like it might be OK,” Theodora thinks out loud, as she looks through her search results. She opens the link to the collection in Google Books, and saves it in Zotero. “I’m going to save it to Zotero. Since I just created that [new collection], I’ll automatically dump it in there,” she explains. Theodora further elaborates that it is the same search process she uses for all electronic publications. The first round is a quick run through search materials, where she collects and saves everything that seems interesting and relevant for the project. In the second round, she looks closely at the saved materials and decides which of them to keep. Finally, she organizes kept files into appropriate folders.

Such a process of collecting and managing electronic information we observed with almost all the humanists who participated in both phases of this project. Faced with a deluge of information available online, humanists commonly collect everything that seems relevant, “dumping” it into their hard drives, citation managers, or Dropbox folders for further examination. Detecting such a trend, Pampson (2014) posits that scholars “unsystematically gather materials, never certain when an object might come in handy,” and observes that over time “folders seem to grow in number and size geometrically” generating a ceaseless flow of information [Papson 2013, 387]. Our study implies, however, that seemingly unsystematic scholars’ information seeking and storing behavior actually has deeper and more systematic roots. One origin of such information behavior relates to the essence of humanistic inquiry, which does not necessarily constitute a search for evidence, but for multiple and often opposing points of view. Humanists thus search for and collect materials until they reach the point of data saturation, as one of our study participants explained:

What I try to do in my criticism in historical writing is to be as much as possible dialogic, to look for competing points of view, and there’s a point at which they [collected materials] become redundant, and then you know you’re probably going as far as you need to go in that direction, so it’s kind of redundancy that begins to tell you [when to end your search].

Example 1. Peter, professor of rhetoric

Peter imports to Zotero all publications that look relevant for his research, but acknowledges that it can take months before he moves from this first round to the second round, when he goes back to Zotero and actually reads those
Audrey also gives the following example: “It’s conceivable, I’d never would come back [to Zotero] and do much for this, but I might,” Peter specifies. He recognizes that this way of working might be inefficient, but stresses that such inefficiencies constitute a critical part of humanistic research. This point is important for broader understanding of humanists’ digital workflows and their interaction with digital technologies. It illustrates that efficiency is not a neutral or an inherent characteristic of digital tools, resources, or workflows independent of scholars’ practices. Rather, efficiency is construed through scholars’ interaction with digital artifacts, as we discuss throughout the paper.

Among the respondents in this study, the only exception from the above-described online search practice was Linda, a research associate in comparative and international literature, whose information behavior included a careful selection of online materials already in the first round of encountering them. Like Theodora, Linda starts her search in Google Scholar by entering a search phrase. Going through the list of results that her search yielded, Linda reads the abstracts and, if an abstract seems relevant, she reads the full text online, taking handwritten notes in her notepad. Linda explains that if she ends up taking a lot of handwritten notes while reading articles online, she later takes photos of those notes, emails the photos to herself, saves them as PDFs, and then OCRs them, thus transforming them into searchable digital notes. When she finds a text that fully meets her criteria and needs, she downloads a copy and saves the PDF file to her Box cloud storage. From there, she moves the file to her library in the Mendeley citation manager, but only if she decides to cite that piece. “I won’t download a PDF until I have significant cause to,” Linda explains. “I have a pretty deep repository [in Mendeley], and these are all things that I’ve read. Most of it I’ve annotated in some way,” she adds. In that way Linda avoids the deluge of downloaded files that stay unused, and keeps her collections of digital materials focused.

This example illustrates how conceptions of efficiency are idiosyncratic and can influence digital scholarly workflows in significant ways. For Theodora and Peter, effectiveness implied keeping the diversity of materials available for further analysis and for serendipitous development of their epistemological paths. For Linda, on the contrary, efficiency implied very selective retention and an early analytical reduction of heterogeneity in her knowledge production process. Such differences might be a matter of personal preferences, and we could say, following Malone’s (1983) seminal classification of storing and sorting practices, that Theodora and Peter belong to the category of *pilers*, who accumulate materials in a loosely organized manner, while Linda belongs to the category of *filers*, who store and organize their materials in a highly systematic way. We could also speculate that such differences may result from dissimilar professional experiences. Both Peter and Theodora are senior scholars, while Linda is a junior one, and one might argue that senior scholars feel more comfortable dealing with heterogeneity of analytical views than their junior colleagues. But one might also argue that junior scholars feel more comfortable dealing with digital tools and resources, and that the observed difference in handling electronic materials derives from that disparity. A look at humanists’ offline search practices supports such a possibility.

While Linda’s search behavior stands as rare among the study participants when it comes to online search, it closely resembles the traditional archival search behavior humanists exhibited prior to the introduction of digital technologies into archival work. Our respondents recollect that, when collecting primary materials in physical archives, they used to carefully read and identify materials most pertinent to their research, systematically storing only a handful of selected documents. The rest of important materials they would summarize on site in written notes, which ranged from short reminders to detailed document digests. Such a practice of managing materials recalls the time when photocopying was one of the main ways of saving archival documents for offsite use. It was neither cheap nor easy to handle piles of paper, so, instead of acting as *pilers* who gather everything, researchers needed to act as *filers*, and to strategically decide which materials were most important to collect.

Audrey explains, for instance, that she used to take extensive notes in her past archival work, and that those notes have been very useful for her over the years. “Now that they let you take pictures in the archives, it’s wonderful, but at the same time you don’t take the time to take notes anymore,” this scholar observes. In talking about her past archival work, Audrey also gives the following example:

A lot of the archives for my field used to be very badly organized, almost like what people told me about being in Poland in 1970 - there was a shop that had opened and there’s a queue and you just
queue up and hope that something will be for sale, whether you need it or not. It was very much like that with archives. They gave you something. You would take furious notes and you would also photocopy, but you might only use one or two things. You put it aside and you think, “I’m always going to be able to use this.”

Audrey explains, however, that all those materials collected in an unintended way are “in a sense all active,” as they informed her later studies, which she did not know at the time of collecting them. In other words, archival search helped Audrey in finding materials she was looking for, but it also served the function of making her aware of other valuable information. Malone (Op.cit.) identifies these two functions as finding and reminding, explaining that a distinction between them rests on intentionality. Specifically, he posits that if a person becomes aware of the information she was looking for, then the finding function has been performed. If, while conducting this search, the person becomes aware of something she was not looking for, yet found useful for her work, then the reminding function has been served [Malone 1983, 106]. This distinction, along with Audrey’s explanation and experience, resonates with Peter’s previously cited description of his “inefficient” search process, demonstrating again that humanists’ outwardly inefficient digital workflows often have deeper epistemological roots and constitute the core of humanistic inquiry.

The above-described archival practice of photocopying selected materials and “taking furious notes” of everything else changed significantly when archives started allowing researchers to take digital photos of materials in their collections. This change overwhelmingly transformed scholars into piles of hundreds, sometimes even thousands of materials, significantly disrupting their workflows, as we describe and discuss in more detail in the following section.

**Storing and organizing research data and materials**

Both phases of our study showed that humanists largely use internal computer hard drives and cloud-based services, such as Dropbox and Box, to store research data and materials, while only a handful use the Penn State institutional repository, ScholarSphere; this finding is consistent with other studies of institutional repositories use (see: [Davis and Connolly 2007]; [Cullen and Chawner 2010]. The most common way in which our respondents organize stored research materials is according to the research project. A typical project folder, whether on a local hard-drive or in Dropbox, is an over-reaching compartment that breaks down into two main sub-folders, one for primary and one for secondary sources. Primary sources mainly consist of archival documents captured as JPEG files, and they are commonly titled according to the archive and/or archival box from which they originated. Secondary sources, such as PDF files of journal articles and books, are stored in a different subfolder. These subfolders further branch into sub-subfolders organized around the subject they cover (a specific topic, period of time, person), the source from which they originated (a physical archive, colleagues, other projects), or in line with the project/workflow phase to which they are related (“phase I,” “articles to read,” “annotated”).

In one of the contextual inquiry sessions we observed, Fred, a historian of science, was organizing and handling materials related to his project on the history of Unidentified Flying Object (UFO) research. The organizational structure of his project materials was the following:

1. the overall project folder titled “UFO project”;
2. a subfolder containing primary sources — thousands of digital photos of archival materials he photographed in various archives. Those materials are further organized into sub-subfolders classified and named after the corresponding archival boxes from which they came, and/or according to a specific research topic they cover;
3. a subfolder containing secondary materials, such as PDF files of articles related to the UFO research. Those materials are classified into sub-subfolders and named according to a decade when they were written — 1950s, 1960s, and so on.

In naming classification units such as folders and subfolders, Fred’s practice mostly meets recommended file-naming conventions; he uses short and descriptive titles, includes the project name in the file name, and so on. But the situation changes when it comes to individual JPEG files of archival materials. Those files all retain numeric, automatically generated file names, without indications about the content of the file. Such a practice mainly stems from the quantity of
JPEG files Fred is managing. He points out that ever since digital photographing became allowed in the archives, he decided to “get the scale of the digitization,” so he took nearly 4,000 photos in one archive alone. Renaming thousands of JPEG files would require substantial time and effort, including more careful reading of the archival documents, so that assigned file-names could reflect their content and enable more efficient browsing and retrieval. Fred, however, quickly reads archival materials while on site and photographs everything that seems relevant for his research. Acting as a *piler*, Fred thus amasses thousands of digital JPEG files that are difficult to organize, rename, search, and retrieve.

Fred first organizes all of the digital materials on his desktop, saving them both on the internal hard drive and in a Dropbox folder. At the end of the day he transfers thus organized materials to an external hard disk. Such a systematic practice of daily backup was rare among the humanities participating in both phases of our study, who generally reported important data loss (see: [Antonijević and Cahoy 2014]; [Antonijević 2015]). Fred stresses that he had suffered a critical system failure and loss of digital materials a few years ago, so he switched to a daily backup routine ever since.

Another professor of history, Connor, describes an experience similar to Fred’s, which however quickly developed into a notably different workflow in managing digital archival materials. When digital photographing became permissible in archives, Connor used to be a *piler* like Fred and to take photos of everything that seemed relevant, saving those materials as JPEG files. After a while, he realized that he would return from the archive with an overload of digital photos that required a lot of storage space, were unsearchable, and for which he often could not recall contextual meaning. Connor thus decided to change his archival workflow. Instead of just snapping photos, he now first reads archival materials at the level needed to determine whether they represent those critical materials that should be saved in the digital form. Resembling Linda’s strategy described in the previous section — to save only those secondary sources that she will cite — Connor comparably stores as JPEG files only those materials that he deems as something he will use and cite in his work. All other relevant archival materials he summarizes on site in short notes, usually a paragraph long, and types as Word files. “This condenses everything really, really small,” Connor says of his work process, commending the ability to recognize key documents that should be stored, and to efficiently summarize all other relevant materials:

> What I have here, for all of that [archival] material, are six pages. It doesn’t exhaust everything. But when I go back to this stuff, I tell myself that former me, back in the fall [when he was doing research in the archive] was smart. I knew what I was doing. I knew exactly what kinds of things were critical and important and what things weren’t.

The observed differences between Connor and Fred’s archival practices bring us back to the question of what constitutes efficiency in humanists’ digital workflows. Like Connor, Fred also recognizes that JPEG files of archival materials do not allow for easy browsing, retrieval, and annotation, but points out that this inefficiency actually benefits his work. It compels him to describe both the data and his ideas about the data in consolidated and focused summaries that he types as Word documents:

> The only way I’ve annotated them [JPEG files] is in this separate [Word] file here. As I would go through [JPEG files], I would keep a paper where I would take my notes on it with ideas. These are going to be things that I want to come back to, themes that I’m noticing in the digital material. So when I want to start to research, I’ll look at this stuff here to tell myself what is it I’m looking for, what is it I want to do, and that becomes a major theses. [11]

For Connor, efficiency means promptly identifying key archival documents that should be photographed and stored, thus eliminating the encumbrance of digital deluge. For Fred, in contrast, efficiency means storing thousands of JPEG files and systematically going through them throughout the research process, slowly discerning over time the main findings and research ideas that those materials suggest, while writing notes that facilitate such continuous analytical engagement with the data.

Using a method similar to Fred’s, Ellen, a historian of American literature, stores hundreds of digitized archival materials and keeps extensive notes that help her navigate through those materials. When working in physical archives, Ellen
uses her cell phone to take photos of archival materials, which she then stores in cloud-based services for further use. “Here I have a folder in Dropbox that says ‘archives,’ and here is the folder for Emory,” Ellen says, explaining that citation managers, such as Zotero, do not handle archival materials well, so she “lives by Dropbox.” This kind of living, however, is haunted by the specter of data deluge. Ellen discloses, for instance, that her work at the Emory archive yielded 189 JPEG files, which she still has not managed to process. “It comes back on your phone, it’s all mixed in with your kids’ baby pictures,” Ellen complains. She specifies that her next step will be to turn those JPEG files into PDFs for each archival object, and to name the PDF files in a useful way that will enable her to browse and retrieve them more efficiently. “That’s how I organize my archival stuff, but it takes a lot of steps to get there, so I have a huge backlog of stuff like that from archives that I just haven’t dealt with,” Ellen relates. Here we see yet another method of managing digitized archival materials, by converting JPEG into PDF files. Neither Ellen nor our other respondents use optical character recognition (OCR) to convert their digitized archival materials into machine-readable text, which illustrates their lack of technological savvy, discussed in more detail in our previous work (see: [Antonijević 2015]).

Ellen’s work sometimes also includes materials available in digital archives, and in such cases, when digital and traditional practices encounter each other, deciding where to store materials and how to organize them is not always a straightforward task. As an example, in a contextual inquiry session we observed Ellen was dealing with an archival document that, according to her established organizational logic, should be stored in the subfolder of primary documents. Yet, Ellen decides to store it differently, in the subfolder for secondary sources. She explains that she does not consider the document “an archival source exactly” because she found it in the digital archive. “I didn’t go work on a collection there [in the physical archive]; I’m just pulling it [from the web], so for me that’s a resource, not an archival document,” Ellen clarifies. This example illustrates an interesting situation in which ontology of research objects cannot be separated from the technological context in which they are produced, in line with Hottois’ (1984) original understanding of technoscience. Authors such as Sinn and Soares (2014) thus observe that humanists prefer original over digitized sources, and think about digital materials as a way of duplicating the originals rather than as a way of replacing them.

Jordan, a professor of history, seconds the words and workflows of both Ellen and Fred. Jordan takes photos and scans of archival materials from physical collections and stores those digital files on his computer. In organizing research materials, he has a folder for every project, and then subfolders and sub-subfolders for every subject. Jordan discloses that all those valuable and meticulously collected materials live on his computer hard-drive and nowhere else. He backups his materials when he “gets into a panic,” which happens approximately every three months. While we observe his work in a contextual inquiry session, Jordan shows us a Word document that he keeps as the main directory of his primary sources. This Word document lists all the materials he had photographed or scanned in various archives, and it helps him locate those materials on the hard drive. “For instance, I have an item in here from the archive in Germany,” Jordan says, pointing at a folder on his computer. “And then I explain here what it is that I have,” Jordan clarifies further, indicating a paragraph in his Word document. “It’s all not terribly systematic, but it’s my system and I know where everything is,” he concludes. Jordan takes extensive notes on the archival materials he is working with, indicating what those materials contain, from which archive they come, with what research themes and/or questions are they connected, and so on.

Jordan is one of the rare scholars consulted in our study who does not use cloud storage in his work. Our respondents overwhelmingly select general storage systems, such as Dropbox, over specialized tools like citation managers. One of the main reasons is that Dropbox enables them to arrange primary and secondary sources side by side, organizing them in an autonomous and personal way — without imposed classification structures — and to navigate through those materials in a way that is both coherent and personal. For scholars, storing and organizing their research materials is not a technical or an information science activity. It is a deeply personal intellectual pursuit of organizing their thoughts in a way that supports a subjective process of knowledge production, and enables them to construct digital workflows as bricolage, from a diverse range of resources and tools, as we discuss in the concluding section of this paper.

Discussion

Both phases of the Digital Scholarly Workflow study indicate that digital technologies have become prevalent and
transformative in humanists’ research, challenging the view of digital scholarship as vital in the sciences, but not in the humanities. Through an ethnographic approach and an analytical focus on contextualized research practices, our study documented myriad ways in which humanists actively interact with digital technologies.

Contextual inquiry conducted in the second phase confirmed the predominance of commercial and crowdsourced resources in the discovery search, recognized in phase I of the study, and showed that humanists mainly use online services to find and access secondary sources, while their search for primary sources largely takes place offline, in the archival collections.

Despite significant digitization efforts over the past few decades, archives and other physical collections remain pivotal for humanists when it comes to finding primary sources. Sinn and Soares (Op.cit.) observe the same trend by looking at citations of digital archival collections compared with other sources. They conclude that although citation of digital sources increases over time, “the impact of digital archival collections when compared with other source materials is not significant when judging by the frequency of citations” [Sinn and Soares 2014, 1798].

Yet, while archives remain a vital site of humanists’ research, archival practices significantly transform with the use of digital technologies. As we observed throughout this paper, the main transformation emerges from the use of digital cameras and humanists’ on-site digitization of archival materials. All of our respondents whose research includes archival work reported taking digital photos of primary sources and storing them for further use. Rutner and Schonfeld (2012) identified the same trend among their respondents, arguing that the widespread use of digital cameras in archives “is perhaps the single most significant shift in research practices among historians, and one with as-yet largely unrecognized implications for the work of historical research and its support” [Rutner and Schonfeld 2012, 11].

Through such practices humanists build rich personal collections of digitized primary sources, which become one of the central information resources in their work. Scholars consulted in both phases of our study stressed that such personal digital collections provided them with desired work mobility and an uninterrupted access to research data and materials. In parallel, however, such collections confronted them with the problem of data deluge, for which our respondents rarely had efficient management strategies.

The problem of data deluge appeared as a consistent challenge with both primary and secondary sources. As discussed in the previous section, humanists often download from the web and save all secondary sources that seem relevant for their work, but it sometimes takes months before they actually return to those materials. By simply dumping downloaded files into hard-drive or cloud storage folders, scholars often forget about those materials and end up investing time in the same search process again. A further consequence is that their already cluttered storage folders become even messier, as they save duplicate copies of files downloaded in the previous search. Some of our respondents thus point out that they had a complete breakdown of organizational system with a shift away from paper copies, and needed to adjust their organizational system to digital sources. An exception are scholars like Linda, who carefully review research materials before deciding to download and store them, as well as scholars like Theodora, who use citation managers that detect duplicate files.

When it comes to primary sources, humanists traditionally managed archival materials by collecting and storing a number of most important documents, while they summarized other relevant materials in extensive research notes. This established method of managing primary sources has been disrupted with the adoption of digital cameras, as we discussed in previous sections. It appears that, as a result of this change, humanists’ online practices have been transposed to the sphere of physical archives, bringing hastier digital workflow to traditionally more methodical archival work. Humanists now commonly handle archival materials in the same way as they handle online sources — quickly collect everything that seems relevant, store it, and examine later.

Humanists’ archival work thus develops into “collection missions,” as Rutner and Schonfeld illustratively dub it, where scholars spend most of onsite time photographing archival materials [Rutner and Schonfeld 2012, 12]. As a result, they regularly return from archives with hundreds, in some cases even thousands of digitally captured materials, which confront them with challenges of storing, searching, annotating, and retrieving those materials. Scholars’ personal digital collections thus need to be a more visible point of analytical attention, particularly with regard to securely storing those
materials and managing them in ways that enable easy retrieval and better workflow integration across tools and data formats.

Namely, the humanists in our study commonly use more than one software solution in their research workflow. Trace and Karadkar (2016) noted similar practices among the humanists they studied, and indicated support for a more integrated solution with fewer moving parts and more connected activities. Literature on personal information management has also discussed the fragmentation problem within an individual workflow ([Al-Omar and Cox 2013]; [Bergman, Beyth-marom, and Nachmias 2007]; [Jones and Anderson 2011]). This “fragmentation of information into different collections forces a person who is working on a single project to store and retrieve items from different locations with no structural connection between them” [Bergman, Beyth-marom, and Nachmias 2007]. In this study, we saw that all of the respondents search in one tool, then deposit in another (or in multiple cases, more than one place), and this disconnect between search and storage is ripe with possible complications, including loss of information. As mentioned in the Introduction, humanists consulted in our study suffered higher data loss than their colleagues in the sciences.

The importance of Microsoft Word as a storage and annotation tool emerged among our participants in that regard. Word is used to store (and then cut and paste into documents) citations and summaries of archival materials; Connor, Fred, and Ellen demonstrate this workflow in action. All of these scholars used Word in isolation as a “blank slate” document where they could create lists or libraries of information that best suit their individual needs.

With this adoption of Microsoft Word in mind, we wonder if it would be useful to build connections into Word that bring other services to the user in this environment (such as article discovery or more facilitated and anticipatory citation management). A number of projects have similarly focused on bringing increased information services into the Word processing environment ([Carr et al. 2004]; [Dickinson and Sefton 2009]; [Fernicola 2009]; [Green and Awre 2009]; [Murray-Rust and Rzepa 2004]; [Sefton et al. 2009]). These initiatives, all conducted between 2000-2010, saw the value of semantic data within Word documents for researchers. An overall goal of most of these projects was simply to combine services and collaboration to provide greater opportunities for researchers. While the projects have not continued, the need persists for better integration of a range of services within the word processing environment, based on empirically observed research practices.

This understanding of research practices mandates that future DH tools designed for the scholarly workflow must account for humanists’ empirically observed needs, which are sometimes slow to change. For instance, we recently worked with researchers at DePaul University to replicate our 2012 Digital Scholarly Workflow study among their faculty. Five years after our original study, the findings were strikingly similar. Humanists continue to struggle with inefficient digital workflows, primarily with tools that have been in use for a decade or more, and which do not successfully support seamless workflows across information sources.

Understanding similarities and differences in humanists’ interaction with diverse information sources — digital and analog, institutional and crowdsourced, academic and commercial, public and personal — will facilitate the development of digital tools and methods that support uninterrupted search, storing, and management across sources, computer programs, and data formats. Such tools are needed to support a seamless manner in which humanists combine research activities spanning a variety of digital and analog sources and tools.[4] In this study, we saw Audrey who started her search by using the Word search function to query a folder with her notes, switched to online search in the Penn State library portal, then moved from the digital to the analog realm reading through her books, and finally resolved her search effort through the use of Google and Wikipedia. We also observed Linda who takes handwritten notes in her notepad while reading articles online, then takes photos of those notes and emails the photos to herself, saves them as PDFs, and finally OCRs them, thus transforming her handwritten notes into searchable digital notes.

These examples show that in humanists’ workflows digital and analog tools and resources do not necessarily replace one another or vie for the dominant position. They rather supplement each other in an interconnected, hybrid, and bricolage manner that comes to our respondents as an organic manifestation of their work. To understand this better, it is useful to return to the view of scholars as bricoleurs, which we presented in the Introduction of this essay, and to
which we return in the next and concluding section of this essay.

**Conclusion**

The view of scholars as bricoleurs draws on Lévi-Strauss’s (Op.cit.) inquiry into bricolage as situated practice and its specific, unscripted approach to technology illustrated in the comparison between engineers and bricoleurs. Although the figures of bricoleur and engineer are idealized — as is the difference between them — they point to an important difference in engaging with technology and knowledge production. While the engineer uses specialized tools adjusting tasks to the availability of those tools and striving for perfection, the bricoleur takes the opposite route and adjusts non-specialized tools to a variety of tasks, tolerating imperfection. The bricoleur achieves that by combining available tools and resources into an improvised aggregate adjusted to his or her needs.

Bricolage is thus a creative act of recombining elements — tools, genes, cultural forms, and so on — in individualized and unscripted ways. Drawing on Lévi-Strauss, Jacob (1977) points out that “to create is to recombine,” and offers the example of natural selection:

> The action of natural selection has often been compared to that of engineers. This, however, does not seem to be a suitable comparison. First, because in contrast to what occurs in evolution, the engineer works according to a preconceived plan. … Second, because… to make a new product he [the engineer] has at his disposal both material specially prepared to that end and machines designed solely for that task. Finally, because the objects produced by the engineer, at least by the good engineer, approach the level of perfection made possible by the technology of the time. In contrast, evolution is far from perfection. … It works like a tinkerer — a tinkerer who does not know exactly what he is going to produce but uses whatever he finds around him. … The tinkerer gives his materials unexpected functions to produce a new object. From an old bicycle wheel, he makes a roulette; from a broken chair the cabinet of a radio. Similarly evolution makes a wing from a leg or a part of an ear from a piece of jaw. [Jacob 1977, 1163–64]

Turkle and Papert (1990) similarly draw on Lévi-Strauss to talk about “hard” and “soft” approaches to programming as two ways of engaging with technology and knowledge production. They interpret these two approaches as gendered, and, more relevant for our discussion, as differently scripted. While “hard” corresponds to highly scripted and canonical approach, “soft” is eminently creative, non-canonical, and personal. Turkle and Papert see computers as a medium for expression of personal styles. Accordingly, they interpret different styles of engaging with computers as insights into various styles of thought, which they term epistemological pluralism. “Hard and soft are more than different approaches to computation,” the authors explain. “The phrase *epistemological pluralism* (rather than, for example, *computational pluralism*) underscores the generality of the issues. The computer forces general questions about intellectual style to reveal an everyday face” [Turkle and Papert 1990, 166]

To illustrate such revealing of intellectual styles, the authors provide examples of two Harvard students dissatisfied with the way programming is taught and practiced — namely, as a prescribed, highly scripted, and blackboxed activity. One of the students wants instead to code in a more flexible and creative manner and to “manipulate computer language the way she works with words when she writes poems” [Turkle and Papert 1990, 164]. Her colleague, a pianist, similarly discloses that although she had learned the required — scripted and canonical — approach to programming, she enjoys programming in a more creative way, treating it as musical phrases [Turkle and Papert 1990, 164].

Humanists consulted in this study similarly engage with digital tools and resources in a non-canonical, highly personal way, which corresponds to their style of thinking and working, to their intimate epistemology. We saw in the previous sections that humanists do not experience their digital workflows as a technical or an information science activity. Instead, it is a subjective process of knowledge production in which they construct digital workflows from a diverse range of tools and resources suited to their preferences, sensibilities, and research needs. As one of us showed in the previous research [Antonijević 2015], humanists’ uptake of digital tools often emerges spontaneously, based on specific research or teaching needs. Once so adopted, digital technologies continue to interact with scholars’ workflows quietly, usually remaining outside the purview of digital humanities. It is, however, critically important for both DH theory and
practice to recognize and engage this variety of voices of non-canonical technology use, which “holds the promise of catalyzing change,” as Turkle and Papert point out [Turkle and Papert 1990, 163].

One of such changes includes recognizing contextualized empirical studies of scholars’ engagement with technology as the foundation for the development of digital research tools. Namely, the bricolage way of working observed in this study often comes from necessity, that is, from the lack of adequate tools and resources. When our respondents use cell phone cameras to digitize archival materials on site, this application of an available and simple technological solution is a response to the lack of specialized tools and digitized primary sources. Similarly, respondents’ extensive Word annotations of archival materials alleviate the pain of managing impromptu digitized sources difficult to search, annotate, and retrieve.

The bricolage way of working thus often stems from the failure to provide digital tools and resources that meaningfully improve scholarly practices and inspire wider uptake. In both phases of the Digital Scholarly Workflow project humanists emphasized that their adoption of digital tools primarily depends on the tool’s capacity to meet their research needs. As long as a simple tool like Word or Dropbox successfully supports their work, scholars do not tend to replace it with a novel or more sophisticated tool. “Sufficient, not efficient” is the guiding motto of their selection and adoption of digital tools and resources.

As we discussed throughout the paper, efficiency cannot be qualified as a neutral or inherent characteristic of digital tools independent of scholars’ practices. It is scholars’ interaction with digital artifacts that determines what constitutes efficiency of a tool in the context of a scholar’s workflow. As Pinch and Bijker (1987) point out, the same technological artifact can be attributed with very different meanings and interpretations depending on the context of use and particular understanding of those artifacts. Law (2003) similarly recognizes that technological imperfections and “unruly” practices in technology use are not only unavoidable, but also beneficial insomuch that they become integrated into a routine that enables uninterrupted functioning.

It is through this analytical lens that we can understand “inefficient” and “unruly” bricolage practices observed in this study. Like bricoleurs in Lévi-Strauss’s original account, humanists are constructing their digital workflow with an orientation on creativity and interpretation rather than on efficiency. Embracing humanities researchers as bricoleurs will thus enable us to prevent the quest for computational efficiency from overshadowing the quest for humanistic understanding.

Notes

[1] In fact, just a brief Google Scholar search reveals that the title of this paper is far less original than we hoped it would be, inadvertently following in the footsteps of Pinar’s (2001) “The Researcher as Bricoleur: The Teacher as Public Intellectual,” and the series that includes student as bricoleur [Schnelker 2006], teacher as bricoleur [Reilly 2011], and so forth.

[2] Three sessions took place at the University library.

[3] Papers 3, a reference management software created by Mekentosj, has a variety of automated features including “smart” collections, which automatically sorts new citations and places them in corresponding collections directed by parameters preset by the user.

[4] Tools such as Tropy (https://tropy.org/) are a positive step in that direction, though they have a lot of space for improvement. For instance, it is essential for such tools to have a mobile application allowing scholars to manage photographed archival materials on site, to enable easy upload and synchronization with cloud-based services, to provide detection of duplicate files, and so on.

Works Cited


