2018

Personal Digital Archiving for Journalists: A “Private” Solution to a Public Problem

Rachel King
rachel.king@liu.edu

Follow this and additional works at: https://digitalcommons.liu.edu/brooklyn_libfacpubs
Part of the Archival Science Commons, and the Information Literacy Commons

Recommended Citation
Rachel King, (2018) 'Personal digital archiving for journalists: a "private" solution to a public problem', Library Hi Tech,
https://doi.org/10.1108/LHT-09-2017-0184

This Article is brought to you for free and open access by the Library at Digital Commons @ LIU. It has been accepted for inclusion in Brooklyn Library Faculty Publications by an authorized administrator of Digital Commons @ LIU. For more information, please contact natalia.tomlin@liu.edu.
Personal Digital Archiving for Journalists: A “Private” Solution to a Public Problem

Purpose: The purpose of this paper is to encourage librarians to teach digital archiving practices to journalists a) as a way of giving journalists the skills they need to save their work for future use and b) to facilitate the preservation of journalism for posterity.

Design/methodology/approach: The author has reviewed the personal digital archiving literature and analyzed how it might be specifically tailored to the unique needs of journalists.

Findings: Daily journalism has traditionally been preserved by libraries in the form of newspapers and magazines housed in library periodicals departments. Now that nearly all journalism is published online and libraries generally only have access via temporary subscriptions, libraries are prevented from doing any kind of traditional preservation work (e.g., storing copies locally). In the future, this lack of local preservation may lead to a shortage of early-21st-century primary source material for historians.

Research limitations/implications: The needs of journalists do vary greatly based on the nature and format of their work and its publication venue, making it difficult to offer a single set of standards or recommendations.

Originality/value: While personal digital archiving advocates have generally interpreted the word “personal” to be synonymous with “private,” this paper points to the need to expand the concept to include professional activities, particularly in light of the
prevalence of telecommuting and freelance work arrangements, and the lack of support and training received by remote workers and independent contractors

Paper type: General review

Keywords: digital journalism, digital news, personal digital archiving, digital preservation, freelancers, independent contractors, telecommuting

Introduction

Increasingly, Americans' most precious information—text, photos, and multimedia—is stored digitally. While this allows computer users to store huge amounts of material while avoiding physical clutter in their homes and offices, digital information has the disadvantage of being fragile: files can easily be lost or damaged. The field of personal digital archiving has grown over the past decade in response to the problem of data loss and the public's relative lack of knowledge about the potential risks to their stored files. Many librarians and archivists have been teaching patrons and stakeholders the ins and outs of protecting their digital files. One striking feature of the personal digital archiving field is the fact that the books, articles, and websites devoted to the topic tend to treat the word "personal" as synonymous with the word "private." Williams et al. (2009) exemplify this tendency when they write, “We adopt the term ‘personal digital archives’ to refer to these informal, diverse, and expanding memory collections created or acquired and accumulated and maintained by individuals in the course of their personal lives, and belonging to them, rather than to their institutions or other places of work.” Many writing on this topic either
unconsciously or consciously define personal digital archives in a similar way. They give advice on the long-term storage for the appurtenances of private life, for example, family photos and legal documents. They put less or no emphasis on items that are professional in nature. However, it is a reality of contemporary American life that freelancers and casual laborers make up a huge part of the workforce. Their “places of work” are located in their homes, or perhaps at nearby libraries or cafés. It has been estimated that more than a third of U.S. workers freelance (Horowitz, 2016). More than 57 million American workers freelanced in 2017, and it has been forecasted that, based on current trends, by the year 2027, the majority of American workers will be freelancers (Freelancers Union and Upwork, 2017). Furthermore, many staffers work from home offices that may not be well supported by a company IT department. According to the American Time Use Survey, 22 percent of Americans work from home at least some of the time; Gallup offers an even higher number, estimating that 37 percent of American workers telecommute (Bureau of Labor Statistics 2017; Jones, 2015). For the modern workforce, then, personal information has more than sentimental value. Catastrophic data loss can have a devastating impact on careers and livelihoods.

One field where this is especially true is journalism. The profession is generally regarded to be in crisis (McChesney and Nichols, 2010). Profit margins are slim and layoffs are common. Furthermore, a sizeable number of professional writers (at least one in six) work freelance (Bureau of Labor Statistics, n.d. c). Comparatively few journalists work full-time for an employer possessing sufficient funds for and a philosophical commitment to expensive and complicated preservation activities (Center for Research Libraries, 2011, Halbert et al., 2016; Skinner and Schultz, 2014). Often,
web-based publications are start-ups operating on shoestring budgets. Furthermore, journalists' needs are very specific: Documents from every part of their process from note-taking to final product have ongoing value to them. Published articles, for example, serve as marketing tools; freelance writers send work samples to editors along with emails in which they pitch story ideas (Ohlheiser, 2017). Old articles and notes can provide trustworthy information that the journalist has already researched extensively. Using them for reference can save a journalist on deadline a great deal of time and money. Bellantoni (n.d.) makes this point, as she describes how she uses her 2008 presidential campaign blog as a reference source: “I constantly refer to it to remember where it was that Obama belted out Aretha Franklin or check the name of an Obama field organizer.”

The present writer has worked as a journalist and freelance writer, and the origin of this paper comes out of a personal experience with lost links. In 2016, Meredith Corporation revamped its More.com website. The site had been created years earlier to support More magazine, a print publication aimed at women over the age of 40. As part of this redesign, the print magazine ceased publishing and the website reinvented itself as an online publication for millennials. The boomer-focused articles were scrubbed from the site and replaced with content designed for a younger demographic. While the public will no doubt survive the loss of this information, the impact on a personal portfolio remains devastating. The present author, for example, lost a bylined article on how patients can make the best medical decisions. Better self-archiving practices would have prevented the loss. While it could be said that on the one hand, this particular article has
little value for posterity, it is also true that this kind of self-effacing attitude—common among writers for hire—makes it difficult to secure freelance journalists’ legacy.

Finally, the ephemeral nature of digital journalists’ work makes it easy for employers to erase anything controversial. Complaints from advertisers or other powerful interests, for example, can result in the removal of articles from websites; Hennelly (2017) describes the disappearance of an article of his from the CBS News MoneyWatch website after Sean Spicer, then a high-level Republican National Committee staffer, complained to the company. Even more dramatically, in November of 2017, the owner of the DNAinfo-Gothamist network of local news websites, Joe Ricketts, abruptly closed the organization—and shut down the sites’ archives at the same time. His swift removal of the archives from the web was widely viewed as an attempt to punish journalists at the New York-based Gothamist site who had voted to establish a union. Until he bowed to public pressure to reinstate the archives a day later, many of the company’s ex-employees found that they were without any access to years’ worth of work. Abby Ohlheiser of the Washington Post described the archives’ disappearance as “a scene from a journalist’s worst nightmare.” One of the New York-based Gothamist reporters quoted in the piece commented, “the last two years of my life have been erased,” (Nolan 2017; Ohlheiser, 2017). (At the time that this journal article went to press, there was more good news to report: In February 2018, a consortium of public radio stations purchased the Gothamist-DNAInfo assets, and it appears that this group, which includes New York City’s WNYC, Southern California’s KPSS, and Washington, D.C.’s WAMU will take on the responsibility of maintaining the archives of Gothamist-DNAInfo’s entire network of local news sites. Still, the long-term fate of Gothamist-DNAInfo assets from cities not
served by these three public radio stations [for example, Chicagoist and San Francisco’s SFist] is a little unclear [WNYC, 2018]).

During the print era, of course, publishers didn’t wield so much power over the archives. Hard copies of publications could not disappear so easily. Articles might be retracted, certainly, but publishers could not pull them from library shelves. Today publishers aren’t the only entities that might try to make an article—or an entire archive—disappear. For example, two-thirds of investigative journalists believe that they have been subjected to digital surveillance, and even more believe that being a journalist increases one’s chances of being a victim of digital spying. Since accounts that are hacked can also easily be tampered with or erased, it is clear that the risks to the long-term survival of digital journalism are great and come from a variety of sources. (Pew Research Center, 2017; Asher-Shapiro, 2018).

**Literature Review**

Data loss is an important problem in digital journalism. A 2014 Missouri School of Journalism survey found that 27 percent of hybrid news organizations and 17 percent of online-only ones have experienced a significant loss of digital content (McCain, 2015). This signals a problem at the organizational level; publications and their parent companies are not allocating resources at a sufficient level to prevent data loss, though it is not clear whether the failure is due to a lack of financial resources or an absence of will. Whatever the reason, journalists who hope that the publications they write for have adequately provided for the long-term preservation of published articles may, based on the picture painted by this study, be disappointed. As Lynch and Fontaine (2014) write,
“As newspapers move increasingly online, our means of archiving them still fail miserably.”

As mentioned previously, articles and books within the field of personal digital archiving tend to focus on private personal information rather than professional personal information. Van House (2011) examines the complex nature of digital preservation as it applies to family photographs, noting, “[d]igital archives require more intentional maintenance. Files have to be migrated from camera memory card to computer, from old computer to new, from computer to external storage. Storage media, file formats and photo-management software become obsolete. Online image hosting sites may fail. Images are deleted.” Marshall (2008) describes the tendency for people to procrastinate rather than to actively work to preserve their personal digital archives, and describes the most common response to the problem as a strategy of "benign neglect."

The field of Personal Digital Archiving has produced a number of helpful guidebooks such as Hawkins’ 2013 *Personal Archiving* and Condron’s 2017 *Managing the Digital You* as well as websites such as [digitalpreservation.gov/personalarchiving/](http://digitalpreservation.gov/personalarchiving/) from the Library of Congress. Although information in these guides can be used by anyone accustomed to creating and storing documents on a personal computer, these works are designed to provide advice for those engaged in activities of private life: organizing digital snapshots, scanning old analog photos, etc.

One possible reason why there has not been more emphasis on the needs of independent contractors or remote workers in the professional literature may simply be because librarians or archivists working in house and on staff at their organizations may
not have an intuitive understanding of how common short-term assignments and remote work are within some professions. For example, according to the Bureau of Labor Statistics “freelance journalist” is a job that epitomizes the contemporary “gig” economy (Wiatrowski, 2016). By contrast, the “work environment” for both librarians and archivists is described as more traditional, and far more likely to require an on-site presence (Bureau of Labor Statistics, n.d. a; Bureau of Labor Statistics n.d. b). Since these changes in Americans’ work lives are being experienced differently within different professions, it may take some time for them to be fully grasped by librarians and archivists studying this emerging field. I happen to understand these issues as they pertain to journalism because of my past experiences as a writer. Understandably, it is very hard for librarians and archivists to understand the professional experiences of people in other fields. This paper is no way intended as a criticism, then, but rather as one attempt to bridge that gap.

There are some areas of the library and archives professional literature that do explore the public uses of private digital archives. Unsurprisingly, these examples tend to involve the library patrons that librarians and archivists working in the academy are likely to come into contact with—students and professors, as well as novelists, poets, and other bellettrists. For example, information professionals understand how important professors’ archives are for ensuring the preservation of the scholarly record. Indeed, academics do much of their work off-campus; scholarly research is both a professional activity conducted in the office or lab, as well as a commitment undertaken year-round at home and on the road. There is evidence that academic librarians have been at the forefront of campus efforts to educate faculty about best practices (Moulaison Sandy et
al., 2017). Furthermore, Krtalić et al. (2016) note that student research should be prioritized as well, writing, “students produce huge amounts of personal digital data and documents. It is therefore important to raise questions about their awareness, responsibility, tendencies, and activities they undertake to preserve their collective digital heritage.” They found that students were interested in preserving their digital information, and that librarians could have a role to play in offering them instruction, encouragement, and direction.

Similarly, the subject of writers’ archives has received some attention. Becker and Nogues’s “Saving-Over, Over-Saving, and the Future Mess of Writers’ Digital Archives: A Survey Report on the Personal Digital Archiving Practices of Emerging Writers” (2012) and Mičunović et al.’s “Literature and Writers in the Digital Age: A Small-Scale Survey of Contemporary Croatian Writers’ Organization and Preservation Practices” (2016) acknowledge the professional needs of working writers. Because writers traditionally work from home offices and have fluid schedules (in contrast to the lives of traditional office workers), their personal computers are also used for professional activities, and there is no clear dividing line between their work time and personal lives.

The authors of these last two papers have what might be called an archivcal or special-collections orientation (Cushing, 2010). Their interest in writers' archives is informed by the need to provide future literary scholars documentation of writers’ workflow, the evolution of drafts, etc. (If one were to describe a generalist librarian’s interest in digital archiving, one might instead emphasize the preservation of and access to the published output.) Both papers are permeated by a romantic sense that lurking
among today's writers are tomorrow's acknowledged literary geniuses. That said, most working writers do not attain iconic status. Many of them are busy creating blogs and social media posts rather than writing novels or publishing poetry manuscripts. In the past, as a matter of course, libraries preserved and provided ongoing access to journalists’ work in published form via periodicals collections. What’s more, they did a less-than-thorough job preserving non-print items such as radio and TV broadcasts. (Indeed, as today’s journalism industry abandons print in favor of a variety of other media, libraries should use this moment of disruption to re-think their relationship to all news media, both print and non-print.)

In 2001, writer Nicholson Baker published *Double Fold*, a book describing what he perceived as libraries’ failure to adequately preserve print newspapers. In this work, he singled out for particular criticism the practice of replacing print newspaper collections with microfilm, as microfilm can only preserve the text contained in a newspaper, while failing to replicate or “preserve” the actual experience of reading a newspaper. Today the problem of preserving both format and content has become exponentially more complex (Center for Research Libraries, 2011). Baker’s print newspaper collection safely resides at Duke University (Duke University Libraries, n.d.). It is hard to imagine how all of today’s multimedia news stories and interactive web features will be preserved given the fact that their presentation is dependent on hardware and software that is rapidly evolving, and some cases becoming obsolete.

**Personal Digital Archiving for Journalists**
Journalists are aware of and concerned about the problem of lost articles and personal archives; some have even published personal accounts. (Beaujon, 2015; Bellantoni, n.d.). The present author, in response to her lost article, created a webinar on personal digital archiving for a national organization of journalists. Librarians who wish to connect with journalists within their local communities can organize in-person events for local chapters, or classes for student journalists. Librarians and archivists can serve the journalism community by offering information about digital archiving, while at the same time serving the greater public by creating conditions that will preserve journalism. Of course, all personal archives have potential value to posterity, but journalists’ archives are uniquely valuable in the way that they document both the events of a time as well as the zeitgeist. Because journalists strive to reach a wide readership, their concerns are not just their own, but reflect the priorities of the entire society (Sinn et al. 2017).

Boss and Broussard (2016) correctly note the difficulty of preserving complex digital news objects such as apps, and the importance of finding ways to save both the information content and format integrity of such objects. Nevertheless, it is also true from the perspective of the individual journalist that imperfect personal solutions (e.g., making screenshots of interactive web features and archiving them for future personal use) offer significant if incomplete benefits. To begin with, the creation of a personal archive will enrich and benefit the individual journalist throughout his or her career. Furthermore, the existence of such personal archives will allow journalists to be stewards of important information throughout their lives, will allow them to appoint digital executors (Carroll and Romano, 2010). The concept that “lots of copies keeps stuff safe” holds true in the digital world, just as it did in the analog world (Frick, 2011). Furthermore, Reyes (2013)
has noted that in order to adequately preserve cultural memory, professional librarians and archivists will need to work with members of the public who have become stewards of personal information.

**The Responsibility of Information Professionals**

This paper is intended to help librarians grasp the extensive preservation needs of many of the populations of patrons the serve. It asks how libraries and archives can be re-imagined in order to creatively address the preservation challenges of different groups. For example, some defining features of contemporary journalism include the dominant role played by multimedia and social media and the low cost of entry. This means both that more people are producing journalistic content on their own for a variety of outlets, and that more journalists are working with a variety of formats. Gone are the days when journalists limited themselves to one medium (for example, either print or broadcast) and when everyone in the field worked for one of a handful of newspapers, magazines, radio or TV stations or networks. The accessibility of digital tools has allowed more people to produce journalistic content on their laptops: writing, photos, video, and audio.

Still, in spite of this daunting explosion in the amount of content, the challenges to preservation continue to be primarily legal, cultural, and financial rather than, strictly speaking technological (Kuny, 1997). That is, the requisite technology exists to preserve these materials. But in order to accomplish the task, money needs to be allocated and legal roadblocks cleared away.
The challenge of making journalists’ archives available on a long-term basis is significant, mainly because of the legal impediments. The literature of personal digital archiving does not tend to grapple with these issues; copyright, which is so pertinent for journalists, is of little relevance for private citizens who are self-archiving privately owned and controlled materials. When individuals are archiving their own snapshots, for example, they hold copyright and may do as they wish. The first sale doctrine—which allows libraries to purchase copies of print newspapers and magazines and do more or less whatever they want with them—facilitated library preservation activities (Reese, 2003). Now, however, librarians may not do what they wish with a digital copy of an article because digital files generally come with a) contracts that prevent copying and b) digital rights management that allows publishers to detect and prevent unwanted uses. Even if a journalist saves a copy of an article she herself has written, her actions are circumscribed, depending on the terms of any contracts she may have signed. Publishers generally retain their copyright to digital articles, even if they are doing little to nothing to preserve ongoing access to them. McCargar et al. (2011) note that newspapers are complex objects from the point of view of copyright, as they are composed of work for hire (e.g. commissioned articles) that they own outright as well as independent content that they don’t own the copyright to (e.g., wire service content). Furthermore, Lynch and Fontaine (2014) demonstrate that over the past few decades, newspapers have begun creating relationships with database vendors such as ProQuest, enabling publishers to monetize the old articles from back issues once relegated newspaper morgues.

This leads to a need to point out one little-discussed way in which the for-profit journalism is failing to meet the needs of the public. That is, it is providing too little
incentive and revenue to encourage preservation, while at the same time tying articles up in an antiquated and complex copyright system. This may mean that publishers may retain copyright in perpetuity to articles that they themselves have lost, and therefore be able to limit the activities of individuals or libraries who have gone to the trouble of saving them. We have seen this dynamic at work in the realm of film preservation, in which the archive that restores a film or an organization that pays for a restoration must defer to the copyright holders, making it impossible for them to, say, charge admission to screenings of the film they have rescued as a way of offsetting the costs of the restoration (Gracy, 2009).

**The Dark Archive**

The conflict between copyright and preservation is not a new one, then. Projects in the realm of academic scholarly publishing such as CLOCKSS and Portico (Mering, 2015) have used a “dark archive model.” Institutional-level solutions will require librarians and archivists to reach out to the journalistic community and co-operate with them on archiving articles for purely preservation purposes—that is, either no access to the articles would be provided or would only be provided only in unusual circumstances. The most extreme version of the dark archive model would dictate that nothing could be accessed except in the event of catastrophic data loss that destroyed the news organizations’ own archives. A less extreme version might make provisions for in-library use by researchers.

These ideas are promising, and are made even more promising by the possibility that in the future universities and non-profits will have deeper involvement in the
production of journalism. Universities’ need for information, and the high cost of library database subscriptions have made them receptive to new publishing models. Their involvement in the world of journalism may lead to new models for funding not just the production and dissemination of journalism, but also its preservation. Still, all solutions are incomplete and provisional, due to the protean and unstable quality of digital information, its dependence on ever-changing hardware and software.

The ongoing preservation of journalism will require working with organizations outside the library. Furthermore, any preservation strategy ought to include outreach to individual journalists. One outcome of the crisis in the journalistic business model and the explosion of DIY technologies is that more and more journalists are disseminating information via blogs and social media, and focusing solely on traditional mainstream publications not creates a risk of missing much of the journalistic landscape. Librarians then, in addition to advocating for the creation of digital repositories, will need to include outreach (to independent journalists), instruction (of personal archiving), and curation (of journalistic work from outside the mainstream) to their skill sets.

Still, the crisis in the preservation of digital news is in some sense the crisis of all digital information. In 1997, Kuny predicted that the problem of digital preservation was so thorny that it could not be solved exclusively at the level of large organizations. Kuny said the solution needed to occur at the level of the “desktop,” too. Helping everyone become better stewards of his or her own digital content, and finding long-term storage for personal archives is a necessary addition to preservation strategies that have previously been focused on the institutional level.
Conclusion

The library and archival community has traditionally been involved in the preservation of journalism. In order to continue in this role in the future, however, they will need to dramatically change their approach to include more outreach and advocacy. They will need to educate journalists about the importance of digital archiving, and provide the requisite training to individuals and organizations. They will also need to continue to call for copyright reform. Finally, libraries and archives might need to accept imperfect solutions in the present (such as dark archives that do not meet librarians’ or the public’s preference for unfettered access) in pursuit of preservation goals that might cause inconvenience now, but will be invaluable for generations to come.

In February of 2018, the Metropolitan New York Library Council announced that it has received a $142,000 grant from the Mellon Foundation to help teach personal digital archiving to independent podcasters. While assisting podcasters is certainly a step in the right direction, more needs to be done with many different populations of freelancers both inside and outside of the field of journalism. Influential national organizations such as ACRL and the Library of Congress might consider amending their official guides and documents in an effort to bring personal digital archiving to more people in a greater variety of professional contexts.

References


