



DE GRUYTER

Preservation, Digital
Technology & Culture

Personal Digital Archiving and Catastrophic Data Loss: Is It Getting Better?

Journal:	<i>Preservation, Digital Technology & Culture</i>
Manuscript ID	PDTC.2022.0020.R1
Manuscript Type:	Article
Keywords:	personal digital archiving, data loss, qualitative content analysis
Abstract:	Given the changing landscape of personal computing, this study asks: 1) has the occurrence of personal catastrophic data loss improved with the maturity of digital technology and services, and 2) have the reasons for data loss changed with the maturing of digital technology? These questions are addressed through a qualitative content analysis of N=136 participant responses from 2013 to 2022 to a question about their personal experience of data loss. The study finds that there is no decline over time in participant’s experience of data loss, and the reasons for data loss continue to be varied. Losing digital information endures as an unpleasant experience for today’s computing users, and personal digital archiving ought to be introduced to young people early on to avoid learning about data loss “the hard way.”

SCHOLARONE™
Manuscripts

Personal Digital Archiving and Catastrophic Data Loss: Is It Getting Better?

Introduction

Growing up my family was really into digital photography and stored everything on our sole computer. Unfortunately, the time came when that Windows computer died, and with it all of our images taken up until 2010. It was pretty devastating. To think now that colossal digital erasure could have been avoided if we had known to back up the photographs to an external hard drive sounds so obvious now, but we had figured technology and our digital images were saved indefinitely. Today I still feel the sting of that avoidable loss.

– 28-year-old participant with female pronouns from 2021

I have been having conversations with students about their personal digital information management and archiving since 2010, including discussions of the “horror stories” of when things go wrong. Stories such as the one above are not uncommon as well as the feelings associated with the experience (e.g., “the sting of that avoidable loss”). These types of personal experiences can be tied to the larger narrative of digital technology, with personal computing diffusing into homes since the 1980s driven by a capitalist logic—first with early adopters and then more broadly—and the technology itself having a mixed reliability and longevity. Since that time, the technology has become *seemingly* more reliable, more portable, and more affordable.

Given the changing landscape of personal computing—from a desktop-based to a hybrid environment of both stationary and increasingly mobile technology supported by internet-based services—this study is interested in whether the increased maturity of digital technology is reducing the experience of catastrophic personal data loss. Are today’s users of digital technology still having experiences of data loss accompanied with negative emotions? To address this topic, the following research questions are posed:

RQ #1 – Has the occurrence of personal catastrophic data loss improved with the maturity of digital technology and services?

RQ #2 - Have the reasons for data loss changed with the maturing of digital technology?

To address these questions, we will use qualitative content analysis on online discussions of $N=136$ participants of their experience of personal digital archiving from 2013 to 2022 to uncover the occurrence of personal catastrophic data loss and the reasons for them, with an objective of understanding what changes have occurred in these narratives over this nearly 10-year period.

Before a more thorough discussion of relevant literature, it is first necessary to define terms. The first term used here is “personal catastrophic data loss.” This term is used to indicate an extensive loss of information or data with an accompanying negative emotion of the person who owns or keeps the data, such as panic, fear, disappointment, anger, feelings of loss or even grief. This term merges notions of “catastrophic loss” and “data loss,” with some additional emphasis on the effect on the individual owner or keeper. “Catastrophic loss” is a term used in the insurance industry to indicate a “a severe event that results in losses that are larger than usual,” such as a hurricane (Bureau of Economic Analysis). “Data loss” is an error condition in which information is destroyed by failure or neglect in the storage, transmission, or processing of digital information (Wikipedia). With the term “catastrophic personal data loss,” we are not looking to create objective data loss conditions that must be met to qualify as such. For example, some extensive data loss may not be catastrophic if the owner or keeper intended to delete it anyway and feels no negative feelings towards its destruction. For catastrophic personal data loss to occur, the data lost can be from years of work, or even eras of one’s life, to simply days of important activity. However, it should be accompanied by feelings of loss from its destruction to qualify as such. Additional context can be found in the next section’s review of relevant literature.

Literature Review

1
2
3 This study is situated within the research and activity in the area of personal digital archiving, which is defined as
4 the “collection, management, and preservation of personal and family materials created in digital media” (Gunn, p.
5 xi). Over the last twenty years, there has been steady research and activity in the area of personal digital archiving.
6 Books include *I, Digital: Personal Collections in the Digital Era* (Lee) and *The Complete Guide to Personal Digital*
7 *Archiving* (Marshall), journal articles in publications like *American Archivist*, *Preservation*, *Digital Technology and*
8 *Culture*, among many others. Personal digital archiving is not merely backing up your hard drive. Like with
9 archiving in general, it involves the appraisal or selection of materials for permanent retention, the arrangement or
10 organization of them, and setting up structures for the long-term persistence and access of the information. A major
11 takeaway from the literature is that broadly individuals don’t do a terribly good job of personal digital archiving.
12 There are lapses or outright failures in any or all of the major components of PDA: appraisal, arrangement,
13 description, preservation and access. Cathy Marshall wrote persuasively about how people brought their physical
14 object archiving practice of “benign neglect” to their digital materials, and finds that benign neglect “is a reasonably
15 effective policy for the care of physical assets, but a dangerous strategy for their digital equivalents.” Subsequent
16 research has confirmed the lackadaisical archiving practices of individuals. For example, Becker and Nogues
17 surveyed 110 writers about their personal digital archiving and found that they had “unchecked, unsystematic
18 proliferation of digital files across various storage and access locations... [that] will continue to lead to the loss of
19 important archival material” (p. 509). Post, in his study of the archiving practices of new media artists, found that
20 artists studied were “thinking of preservation on a different time scale than most collecting institutions, maintaining
21 artworks for a few years as opposed to several decades or longer” and they were “on the whole, more interested in
22 sustaining an ongoing creative process than maintaining single artworks” (p. 724). Zhao, Duan and Yang, in their
23 study of the personal digital archiving of postgraduates in Wuhan, China, found that there was a high degree of
24 “technical optimism” and lack of motivation, leaving a segment of students in danger of losing personal digital
25 records (p. 8).

26 Given that individuals show mixed commitment to personal digital archiving, it is worth asking the question if
27 personal digital archiving even matters? While there may be some individuals who are perfectly content and will
28 continue to be for their entire lives without any reminders of the past—be they photos, writings, or other records—the
29 majority of the literature on the subject find that humans find value in these reminders of the past, helping
30 narrativize life and give it meaning. Millar writes that “Records are not memories. Rather, they are triggers or
31 touchstones that lead to the recollection of past events” and they are “...a means by which we gain knowledge of
32 ourselves and our society, leading ultimately, one hopes, to greater understanding, compassion, and wisdom” (p.
33 144, p. 119). In popular culture we often see personal archives employed when one is looking to tell the story of
34 their life in various forms, such as through memoir or documentary. We will read about or see the subject taking
35 paper records and artifacts out of shoeboxes, filing cabinets, or other pre-digital spaces. For example, in the memoir
36 of psychoanalyst and technology researcher Sherry Turkle, she writes about her “memory closet” from childhood,
37 which was a closet that included various “treasures” that she would use to learn about herself and her family (e.g.,
38 family photos, documents that included her father’s name, etc.) (p. 12). Her early engagement with the “memory
39 closet” she ties to her lifelong interest in “objects to think with,” including computers (p. 204).

40 With individuals now assessing their life, with much of the earlier part pre-digital, it is worth asking the question:
41 what will be the equivalent of the younger generation’s “memory closet”? While there may likely still be a shoebox
42 with artifacts in it, there may be less printed photographs, and it is reasonable to expect that many things may be
43 digital, whether they are stored on a computer or in the cloud. Will the younger generation be successful in
44 maintaining a digital “memory closet” over the long-term? If so, what role will social media play in that? Existing
45 research by Coccio indicates that “When social media collapse, youth are disproportionately at risk of losing their
46 digital contributions, because young people disproportionately use the sites that fail” (p. 8). Therefore, assuming the
47 content uploaded to a site like Instagram or Facebook in 2023 will still be accessible in 2063 seems unlikely if we
48 look at the past rise and fall of social media sites, like MySpace and Friendster, the first of which exists in name but
49 contains nothing that it did in the 2000s, or Friendster, which doesn’t exist at all. It is unclear and only time will tell
50 if this tendency to discard past digital efforts will continue.

51
52 The research questions assume that digital technology has indeed matured. This may be a bit of a
53 mischaracterization, as maturity indicates that the technology is at a full or advanced state of development, and that
54 each state is better than the previous. While in some ways the technology has changed for the better, “change” is the
55 unifying thread running through its development. Kirschenbaum writes compellingly about the early adopters of
56 word processors in the literary community emerging in the 1980s. The word processors and their software have
57
58
59
60

names largely forgotten, like Lexitron, Cromemco, Perfect Writer, McWilliams, Wangwriter, Tandy, Osborne, and WordStar, finally giving way to more familiar names like DOS and Macintosh as the decade progressed. With such a profusion of platforms, it is reasonable to assume that maintaining that information in its original format would be difficult if not outright impossible. With the wider adoption of digital technology from the 1990s through 2010s, with two platforms being used nearly ubiquitously by personal computing users (Windows and Macintosh), it is plausible to assume that maintaining a digital “memory closet” may be becoming easier. More recent ways to store large amounts of information, such as the move toward solid-state hard drives that don’t have spinning platters, could potentially aid the longevity of digital information. With the widespread use of cloud computing that manages files, like Google Drive, Dropbox.com, and Apple iCloud, it may appear that the task continues to be simplified, since the management of the computing components is outsourced to IT experts. Or perhaps this is a narrative that the technology industry would like us to believe? This project is particularly interested in seeing if there has been changes over time in users’ experience of their ability to maintain digital information now that technologies like cloud computing are used widely.

Methods

This study will use qualitative content analysis on $N=136$ participant responses to uncover the extent to which there are changes in their experience of data loss using digital technology over time. For example, if digital technology has indeed matured over time and services like cloud computing are helping preserve digital information, we would expect to see a decline in responses that discuss extensive or catastrophic data loss.

These $N=136$ responses were collected from graduate students during the course of a graduate course in Library and Information Science from 2013 to 2022. In this class, students responded to a discussion board post on personal digital archiving and were asked in an open-ended manner, “Have you had any digital preservation challenges in your past experiences (e.g., unable to access digital content) and what were they?” Students are encouraged to write 1-2 paragraphs each week on the discussion board. In some years, I also asked questions about readings related to personal digital archiving, so in some cases students choose to discuss the readings rather than their personal experience. This study will only be looking at responses related to the personal experience students had with digital preservation, and will set aside responses that are outside of personal experience.

These discussion posts were collected from two learning management systems (Moodle and later Canvas), and placed into a single text file. Each individual message was then added to a Google sheet as a row. To maintain the privacy of respondents, the names were deleted from the dataset after some basic participant metadata was added, such as year and semester the class was taken, gender pronouns and year of birth. Then, two graduate students were selected to code the data. One coder was a graduate student in Library and Information Science with interests in the area of archival studies. The second coder was a graduate student in a Master’s program in user experience design without any background knowledge on archives. The researcher discussed the codebook with the coders and asked them to code each message for the experience of data loss reflected in the message, from no data loss to extensive loss of information. They were also asked to code for why the respondent lost information, such as because of theft, lost hardware, failed hardware, service discontinued, damaged/destroyed hardware, among other reasons for loss, including a combination of factors.

Once the data is coded, statistical analysis will take place. This includes a linear regression to determine if the passing of time is related to the severity of data loss, both using the year the course was taken, as well as the year of birth. For example, if extensive data loss was on the decline, we would expect to see a linear relationship with one or both of our time variables. Descriptive statistics will also be produced to shed light on this relationship, and quotes from respondents will be provided to tie the statistics to individuals’ own narratives.

Before getting to the results, we should note that Cohen’s Kappa was run to determine if there was agreement between the two coders, also known as inter-rater reliability. There was fair agreement between the two coder’s judgments, $\kappa = .397$. $p < .001$. For analyzing statistics from the coded discussion posts, we will use coder’s number one’s coding, knowing that it aligns fairly with coder number two’s judgements. Using this coding, 56 responses were excluded from the dataset because they did not discuss personal experience, thus yielding $N=136$ responses.

In terms of demographics of the respondents, the average age of the students is between 29 and 30 years old. 78.4% use she/her/hers pronouns, 17.9% use he/him/his pronouns, and 3.7% use they/them/their pronouns.

Results

Results from the qualitative content analysis indicate that there is no pattern over time related to personal data loss. For example, if we look at the year the course took place and the average experience of personal data loss, with a scale of “1” being no loss and “3” being extensive loss, we see no clear pattern, with the average going up and down from 2013 to 2022, as shown in both Figure 1 and Table 1. Further, a linear regression confirmed that there is no linear relationship between time and experience of data loss, both for the year the course was taken as well as the year the student was born. Thus, the experience of data loss as time goes by is not on the decline.

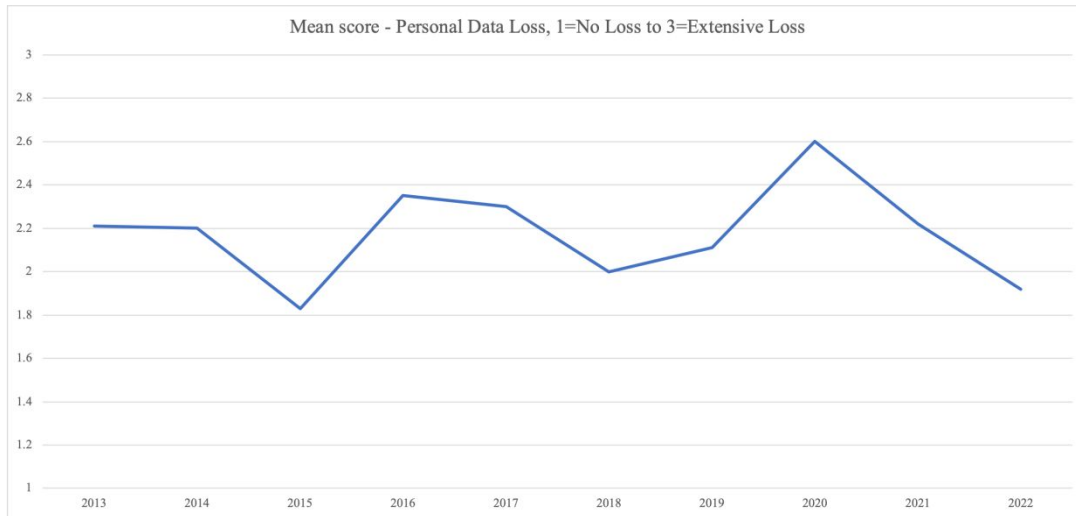


Figure 1. Mean score of personal data loss, which goes up and down over time, indicating no clear pattern

Table 1. Mean score of personal data loss with N respondents and standard deviations.

Year of Class	Mean - Personal Data Loss, 1=No Loss to 3=Extensive Loss	N	Std. Dev.
2013	2.21	28	0.738
2014	2.2	10	0.422
2015	1.83	6	0.753
2016	2.35	17	0.786
2017	2.3	10	0.823
2018	2	6	1.095
2019	2.11	9	0.601
2020	2.6	15	0.632
2021	2.22	23	0.6
2022	1.92	12	0.515

The qualitative narratives provide texture to the statistics, and shed light on the emotionally fraught nature of some of the experiences. For example, a 30-year-old student with they/them/their pronouns writes about a failed laptop in 2020:

I was gifted a laptop before undergrad, and I used it all the time. It had every essay, artwork and important document[s] stored on it's hard drive. The laptop functioned without a problem for 6 years. During that time, I

1
2
3 *never thought to back-up any files, nor did I purchase an external hard drive for the information. Eventually,*
4 *the laptop began over-heating. I decided to invest in a new machine, and left the old laptop hoping that I would*
5 *"get to it" another time. When I finally decided to work on transferring files and data to an external hard drive,*
6 *the old laptop would not even start. It was really disappointing. I lost most of my reference photos for paintings,*
7 *which I had decided later were actually more important than the paintings themselves. Now, I use two external*
8 *terabyte hard drives. So many of my friends who are artists have one back-up external hard drive, but I have*
9 *heard terrible anecdotes [sic] about theft, loss and functional deterioration [sic]. I choose to have two just for*
10 *that reason. Storage seems to be an important part of personal digital archiving.*

11 The above example highlights the negative emotions attached to these experiences of data loss (e.g., “really
12 disappointing”). It also highlights that hardware failure continues to be an issue. While straightforward failed
13 hardware is regularly experienced, often in the form of a failed laptop, the profusion of information across multiple
14 devices and services can also open opportunities for data loss. For example, a 28-year-old respondent with
15 she/her/hers pronouns discusses both loss on a mobile device and on social media in 2020:

17 *My iPhone is always at least two generations out of date, which can cause problems with compatibility, and the*
18 *asymmetrical storage space makes it difficult to keep my music on my phone. Also, like [name removed], I have*
19 *had issues with not being able to move my music library from one computer to another! It can take a long time*
20 *to amass and curate a library like that, and losing all that data is disheartening. / I also encountered a situation*
21 *similar to [name removed]; I thought that my Facebook messages would be stored forever, since, for a period*
22 *of time, I could go all the way back to the beginning of all of my conversations, and they were fully searchable.*
23 *However, about a year or so ago, that stopped being the case. I don't know for sure if they're deleted after a set*
24 *period of time or if the search function only goes back so far.*

26 The above example highlights the negative emotions associated with losing data or being unable to access
27 information (e.g., “losing all that data is disheartening”).

29 Services provided by commercial entities like Facebook are not the only source of data loss. Several students also
30 report losing access to college email accounts upon graduating as a source of lost information. For example, a 28-
31 year-old student with she/her/hers pronouns writes the following in 2020:

33 *Oof, the experiences shared by my peers are painfully familiar. Like [names removed], I was somewhat blasé in*
34 *navigating the transition from my undergraduate time. Despite receiving numerous recommendations from the*
35 *institution to migrate my emails before my account was closed, I allowed my undergraduate email to disappear*
36 *into the ether. I had not recognized how many of my professional and academic contacts were associated only*
37 *with that account until I no longer had access to it. Since, I have been more mindful about which account I use*
38 *for conversations and also have plans to migrate my [institution name removed] account information prior to*
39 *graduation.*

41 In the above example, recognition of a “painful” familiarity summarizes a collective experience of losing important
42 information.

43 Of course, not all narratives included a story of catastrophic data loss. However, they point to other issues such as
44 being overwhelmed with a profusion of information across devices and platforms. For example, the following was
45 offered by a 30-year-old student in 2019 with she/her/hers pronouns:

47 *...currently, I am dealing with the issue of having my digital assets being distributed over many different*
48 *devices, some overlapping, and some completely independent. I obtained a new laptop in the summer of*
49 *2017, unlike some people I did not migrate all of the contents from my old laptop onto the new device, as I*
50 *thought that I would transfer them onto a hard drive. On the old laptop contains coursework from my*
51 *undergrad, other important documents, photos, etc. My new laptop has similarly important documents and*
52 *photos, and some overlap with my old laptop. I also have an external hard drive that is over 10 years old*
53 *and use cloud software, which holds materials from various moments in time. Copies that are distributed*
54 *across different devices are the same, where the resolution, file format, and name remain the same. /*
55 *However, with all my digital assets occupying different platforms, I run into the issue of not knowing where*

specific items reside. The challenge is getting to the material efficiently without having to open 4 different devices/software.

Thus, while some participants may have avoided a terrible data loss, there may be some anxiety around the proliferation of information across multiple devices.

In summary and with respect to research question #1, or has the occurrence of personal catastrophic data loss improved with the maturity of digital technology and services, we can see no evidence in this dataset that this is the case. With respect to the research question #2, or have the reasons for data loss changed with the maturing of digital technology, here we also don't see a clear pattern over time for the reasons for data loss, as indicated in Table 2. For example, if we were to assume that individuals were using cloud computing to maintain their personal information, then we might expect to see a decline in issues related to hardware, or more issues related to "services." This, however, is not the case and we still see "failed hardware" as being an issue reported frequently in both 2021 and 2022. This is also borne out in the qualitative narratives included above. Thus, a variety of factors, and in some cases a combination of factors, leads to personal data loss without any clear changes or trends over time.

Table 2. Reasons for data loss

Reason for Loss / Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
No loss or not discussed	5	0	2	3	2	3	1	0	2	2	20
Unspecified Loss	3	1	0	0	1	1	0	1	0	1	8
Lost Hardware	0	1	0	0	0	0	0	1	0	0	2
Failed Hardware	8	1	1	1	2	0	1	2	7	6	29
Service discontinued	2	3	0	1	0	1	0	0	2	2	11
Damaged/destroyed hardware	0	0	1	0	1	0	1	1	1	0	5
Data/information overwhelm	0	0	0	0	0	0	1	1	1	0	3
Outdated software	1	0	0	0	0	0	0	0	0	1	2
Outdated hardware	1	1	0	0	0	0	1	0	1	0	4
Improper Labeling	0	0	0	0	0	0	0	1	1	0	2
Combination of factors	8	3	2	11	3	1	4	8	8	0	48
Total	28	10	6	16	9	6	9	15	23	12	134

Discussion

This study found that as time has passed (from 2013-2022) that there has not been a decline in individual's experience of data loss. While changes have occurred in the data storage area, such as the widespread use of cloud computing and solid-state storage, this has not led to improvement in the personal experience of data loss. Narratives around data loss still include negative emotions, such as pain and disappointment, indicating that catastrophic data loss continues to persist. The reasons for data loss also continue to vary without any clear trends, such as hardware failure, discontinued service, or a combination of factors.

If the technology itself is indeed maturing, why are we not seeing a decline in the experience of personal data loss? One potential reason is that while the technology itself might be getting better, individuals are running more parts of their lived experiences through digital technology, and for that reason there is simply more information that can be lost, forgotten or mismanaged. Whereas the literary figures Kirshenbaum studies like John Updike and Ralph Ellison may have had some small facet of their life on the computer (e.g., their literary production measured in a few kilobytes per week), today's users are using digital technology to mediate vast segments of their lives. Even with better technology, data loss is still likely to occur because there is so much of it.

One limitation of this study is that the respondents are from a fairly limited demographic: graduate students (above average education), 29-30 years old (younger adults), and 78.4% using she/her/hers pronouns. For this reason, caution should be used in applying these findings to larger and more diverse populations. A further limitation of this study is that it only looked at changes over a time window of nine years (2013-2022), and it may be worthwhile to revisit these experiences of personal data loss over a longer time frame, such as 15 or 20 years.

Conclusion

1
2
3
4 In conclusion, this study found that despite the advances in digital technology, personal computing users over time
5 continue to experience catastrophic personal data loss. While the technology itself may be advancing, because more
6 facets of life are mediated through it, there are simply more opportunities for information to be lost, and thus there is
7 no downward trend in data loss. The reasons for data loss continue to be varied, such as hardware failure and
8 discontinued services.

9
10 While all of the study subjects were college graduates, none of them mentioned learning about personal digital
11 archiving in school, either in secondary school or college. This would indicate that there may be opportunities to
12 expose students to personal digital archiving at a younger age, perhaps through information literacy or computing
13 skills curriculum. Many students in the study learned about data loss “the hard way”—through actually losing their
14 data—and this need not be the case. As this study found that these experiences of data loss are accompanied by
15 negative emotions, such as pain and disappointment, and that these records may be useful in recounting one’s life as
16 a digital “memory closet,” then exposing students to basic curriculum on personal digital archiving may be
17 worthwhile, such as inventorying where important records are located, and if there are multiple copies. Basic
18 education on personal digital archiving can help alleviate “that sting of avoidable loss” and buttress young peoples’
19 future digital memory closets.

20 21 22 **References**

23 Becker, Devin and Collier Nogue. “Saving-Over, Over-Saving, and the Future Mess of Writers' Digital Archives: A
24 Survey Report on the Personal Digital Archiving Practices of Emerging Writers.” *American Archivist* 75.2
25 (Fall/Winter 2012): 482-513.

26
27 Bureau of Economic Analysis. “What is considered a catastrophic loss?” <https://www.bea.gov/help/faq/1354>
28 (accessed 3 June 2022).

29
30 Cociolo, Anthony. “Youth Deleted: Saving Young People’s Histories after Social Media Collapse.” Paper
31 presented at the *International Internet Preservation Consortium General Assembly*, May 19-23, 2014,
32 Paris, France. http://www.thinkingprojects.org/youth_deleted_iipc.pdf (accessed 7/1/2022).

33
34 Gunn, Chelsea. “Introduction: Putting Personal Digital Archives in Context.” In *The Complete Guide to Personal*
35 *Digital Archiving*, edited by Brianna H. Marshall, xi-xxii. Chicago: ALA Editions, 2018.

36
37 Kirschenbaum, Matthew G. *Track Changes: A Literary History of Word Processing*. Cambridge, MA: Belknap
38 Press of Harvard University, 2016.

39
40 Lee, Christopher A. I. *Digital: Personal Collections in the Digital Era*. Chicago: Society of American Archivists,
41 2011.

42
43 Marshall, Brianna H. (Ed). *The Complete Guide to Personal Digital Archiving*. Chicago: ALA Editions, 2018.

44
45 Marshall, Catherine C. “Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field.” *D-Lib*
46 *Magazine* 14.3/4 (March/April 2008). <https://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html>
47 (accessed 7/1/2022).

48
49 Millar, Laura. “Touchstones: Considering the Relationship between Memory and Archives.” *Archivaria* 61
50 (September 2006): 105-126.

51
52 Post, Colin. “Preservation practices of new media artists: Challenges, strategies, and attitudes in the personal
53 management of artworks.” *Journal of Documentation* 73.4 (2017): 716-732.

54
55 Turkle, Sherry. *The Empathy Diaries: A Memoir*. New York: Penguin.

56
57 “Data loss.” Wikipedia article, available at https://en.wikipedia.org/wiki/Data_loss (accessed 6/3/2022).

1
2
3
4 Zhao, yue, Xian'e Duan and Haijuan Yang. "Postgraduates' personal digital archiving practices in China: Problems
5 and strategies." *Journal of Academic Librarianship* 45.5 (2019): 1-11.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only

Personal Digital Archiving and Catastrophic Data Loss: Is It Getting Better?

Introduction

Growing up my family was really into digital photography and stored everything on our sole computer. Unfortunately, the time came when that Windows computer died, and with it all of our images taken up until 2010. It was pretty devastating. To think now that colossal digital erasure could have been avoided if we had known to back up the photographs to an external hard drive sounds so obvious now, but we had figured technology and our digital images were saved indefinitely. Today I still feel the sting of that avoidable loss.

– 28-year-old participant with female pronouns from 2021

I have been having conversations with students about their personal digital information management and archiving since 2010, including discussions of the “horror stories” of when things go wrong. Stories such as the one above are not uncommon as well as the feelings associated with the experience (e.g., “the sting of that avoidable loss”). These types of personal experiences can be tied to the larger narrative of digital technology, with personal computing diffusing into homes since the 1980s driven by a capitalist logic—first with early adopters and then more broadly—and the technology itself having a mixed reliability and longevity. Since that time, the technology has become *seemingly* more reliable, more portable, and more affordable.

Given the changing landscape of personal computing—from a desktop-based to a hybrid environment of both stationary and increasingly mobile technology supported by internet-based services—this study is interested in whether the increased maturity of digital technology is reducing the experience of catastrophic personal data loss. Are today’s users of digital technology still having experiences of data loss accompanied with negative emotions? To address this topic, the following research questions are posed:

RQ #1 – Has the occurrence of personal catastrophic data loss improved with the maturity of digital technology and services?

RQ #2 - Have the reasons for data loss changed with the maturing of digital technology?

To address these questions, ~~we will use~~ qualitative content analysis ~~was completed~~ on online discussions of $N=136$ participants of their experience of personal digital archiving from 2013 to 2022 to uncover the occurrence of personal catastrophic data loss and the reasons for them, with an objective of understanding what changes have occurred in these narratives over this nearly 10-year period.

Before a more thorough discussion of relevant literature, it is first necessary to define terms. The first term used here is “personal catastrophic data loss.” This term is used to indicate an extensive loss of information or data with an accompanying negative emotion of the person who owns or keeps the data, such as panic, fear, disappointment, anger, feelings of loss or even grief. This term merges notions of “catastrophic loss” and “data loss,” with some additional emphasis on the effect on the individual owner or keeper. “Catastrophic loss” is a term used in the insurance industry to indicate a “a severe event that results in losses that are larger than usual,” such as a hurricane (Bureau of Economic Analysis). “Data loss” is an error condition in which information is destroyed by failure or neglect in the storage, transmission, or processing of digital information (Wikipedia). With the term “catastrophic personal data loss,” we are not looking to create objective data loss conditions that must be met to qualify as such. For example, some extensive data loss may not be catastrophic if the owner or keeper intended to delete it anyway and feels no negative feelings towards its destruction. For catastrophic personal data loss to occur, the data lost can be from years of work, or even eras of one’s life, to simply days of important activity. However, it should be accompanied by feelings of loss from its destruction to qualify as such. ~~Additional context can be found in the next section’s review of relevant literature.~~

The research questions assume that digital technology has indeed matured. While time has passed since computing has left mainframe rooms and entered people’s homes and hands, ~~this may be a bit of a mischaracterization, as maturity indicates that the technology is at a full or advanced state of development, and that each state is better than~~

the previous. While we don't know exactly where the advancement of digital technology will lead, and digital technology of the 2020s will likely one day seem "stone age," by "maturing of digital technologies and services" we are largely just pointing out that as time passes the technology changes. For example, Matthew While in some ways the technology has changed for the better, "change" is the unifying thread running through its development. Kirschenbaum writes compellingly about the early adopters of word processors in the literary community emerging in the 1980s. The word processors and their software have names largely forgotten, like Lexitron, Cromemco, Perfect Writer, McWilliams, Wangwriter, Tandy, Osborne, and WordStar, finally giving way to more familiar names like DOS and Macintosh as the decade progressed. Today, we look back at technologies like the Wangwriter and think of it as the early days of word-processing. Thus, because we may still be in the early days of digital technology, when we say "maturing of digital technologies" it is just to indicate that as time passes the technology changes, and if history is a guide then it will continue to do so. With such a profusion of platforms, it is reasonable to assume that maintaining that information in its original format would be difficult if not outright impossible. With the wider adoption of digital technology from the 1990s through 2010s, with two platforms being used nearly ubiquitously by personal computing users (Windows and Macintosh), it is plausible to assume that maintaining a digital "memory closet" may be becoming easier.

More recent ways to store large amounts of information, such as the move toward solid-state hard drives that don't have spinning platters, could potentially aid the longevity of digital information. With the widespread use of cloud computing that manages files, like Google Drive, Dropbox.com, and Apple iCloud, it may appear that the task continues to be simplified, since the management of the computing components is outsourced to IT experts. Or perhaps this is a narrative that the technology industry would like us to believe? This project is particularly interested in seeing if there has been changes over time in users' experience of their ability to maintain digital information now that technologies like cloud computing are used widely.

Literature Review

This study is situated within the research and activity in the area of personal digital archiving, which is defined as the "collection, management, and preservation of personal and family materials created in digital media" (Gunn, p. xi). Over the last twenty years, there has been steady research and activity in the area of personal digital archiving. Books include *I, Digital: Personal Collections in the Digital Era* (Lee) and *The Complete Guide to Personal Digital Archiving* (Marshall), journal articles in publications like *American Archivist*, *Preservation, Digital Technology and Culture*, among many others. Personal digital archiving is not merely backing up your hard drive. Like with archiving in general, it involves the appraisal or selection of materials for permanent retention, the arrangement or organization of them, and setting up structures for the long-term persistence and access of the information. A major takeaway from the literature is that broadly individuals don't do a terribly good job of personal digital archiving. There are lapses or outright failures in any or all of the major components of PDA: appraisal, arrangement, description, preservation and access. Cathy Marshall wrote persuasively about how people brought their physical object archiving practice of "benign neglect" to their digital materials, and finds that benign neglect "is a reasonably effective policy for the care of physical assets, but a dangerous strategy for their digital equivalents." Subsequent research has confirmed the lackadaisical archiving practices of individuals. For example, Becker and Nogues surveyed 110 writers about their personal digital archiving and found that they had "unchecked, unsystematic proliferation of digital files across various storage and access locations... [that] will continue to lead to the loss of important archival material" (p. 509). Post, in his study of the archiving practices of new media artists, found that artists studied were "thinking of preservation on a different time scale than most collecting institutions, maintaining artworks for a few years as opposed to several decades or longer" and they were "on the whole, more interested in sustaining an ongoing creative process than maintaining single artworks" (p. 724). Zhao, Duan and Yang, in their study of the personal digital archiving of postgraduates in Wuhan, China, found that there was a high degree of "technical optimism" and lack of motivation, leaving a segment of students in danger of losing personal digital records (p. 8).

Given that individuals show mixed commitment to personal digital archiving, it is worth asking the question if personal digital archiving even matters? While there may be some individuals who are perfectly content and will continue to be for their entire lives without any reminders of the past—be they photos, writings, or other records—the majority of the literature on the subject find that humans find value in these reminders of the past, helping narrativize life and give it meaning. Millar writes that "Records are not memories. Rather, they are triggers or

1
2
3 touchstones that lead to the recollection of past events” and they are “...a means by which we gain knowledge of
4 ourselves and our society, leading ultimately, one hopes, to greater understanding, compassion, and wisdom” (p.
5 144, p. 119). In popular culture we often see personal archives employed when one is looking to tell the story of
6 their life in various forms, such as through memoir or documentary. We will read about or see the subject taking
7 paper records and artifacts out of shoeboxes, filing cabinets, or other pre-digital spaces. For example, in the memoir
8 of psychoanalyst and technology researcher Sherry Turkle, she writes about her “memory closet” from childhood,
9 which was a closet that included various “treasures” that she would use to learn about herself and her family (e.g.,
10 family photos, documents that included her father’s name, etc.) (p. 12). Her early engagement with the “memory
11 closet” she ties to her lifelong interest in “objects to think with,” including computers (p. 204).

12
13 With individuals now assessing their life, with much of the earlier part pre-digital, it is worth asking the question:
14 what will be the equivalent of the younger generation’s “memory closet”? While there may likely still be a shoebox
15 with artifacts in it, there may be less printed photographs, and it is reasonable to expect that many things may be
16 digital, whether they are stored on a computer or in the cloud. Will the younger generation be successful in
17 maintaining a digital “memory closet” over the long-term? If so, what role will social media play in that? Existing
18 research by Cocciolo indicates that “When social media collapse, youth are disproportionately at risk of losing their
19 digital contributions, because young people disproportionately use the sites that fail” (p. 8). Therefore, assuming the
20 content uploaded to a site like Instagram or Facebook in 2023 will still be accessible in 2063 seems unlikely if we
21 look at the past rise and fall of social media sites, like MySpace and Friendster, the first of which exists in name but
22 contains nothing that it did in the 2000s, or Friendster, which doesn’t exist at all. It is unclear and only time will tell
23 if this tendency to discard past digital efforts will continue.

24
25 Loss of digital information is strongly tied to the means of storing digital information. One of the major
26 developments in the area of data storage that may have the potential to help with preventing data loss is the move
27 toward solid-state hard drives that don't have spinning platters. As devices that have moving parts need lubrication,
28 and lubrication cannot be maintained indefinitely, we may suppose that storage devices with no moving parts may
29 have greater longevity than those with moving parts. The other development in the area of storing digital
30 information is the widespread use of cloud computing that manages files, like Google Drive, Dropbox.com, and
31 Apple iCloud. With these services, the management of the computing components is outsourced to IT experts. This
32 project is particularly interested in seeing if there has been changes over time in users’ experience of their ability to
33 maintain digital information now that technologies like cloud computing and solid-state hard drives are used widely.

34
35 ~~The research questions assume that digital technology has indeed matured. This may be a bit of a~~
36 ~~mischaracterization, as maturity indicates that the technology is at a full or advanced state of development, and that~~
37 ~~each state is better than the previous. While in some ways the technology has changed for the better, “change” is the~~
38 ~~unifying thread running through its development. Kirschenbaum writes compellingly about the early adopters of~~
39 ~~word processors in the literary community emerging in the 1980s. The word processors and their software have~~
40 ~~names largely forgotten, like Lexitron, Cromemco, Perfect Writer, McWilliams, Wangwriter, Tandy, Osborne, and~~
41 ~~WordStar, finally giving way to more familiar names like DOS and Macintosh as the decade progressed. With such~~
42 ~~a profusion of platforms, it is reasonable to assume that maintaining that information in its original format would be~~
43 ~~difficult if not outright impossible. With the wider adoption of digital technology from the 1990s through 2010s,~~
44 ~~with two platforms being used nearly ubiquitously by personal computing users (Windows and Macintosh), it is~~
45 ~~plausible to assume that maintaining a digital “memory closet” may be becoming easier. More recent ways to store~~
46 ~~large amounts of information, such as the move toward solid-state hard drives that don't have spinning platters,~~
47 ~~could potentially aid the longevity of digital information. With the widespread use of cloud computing that manages~~
48 ~~files, like Google Drive, Dropbox.com, and Apple iCloud, it may appear that the task continues to be simplified,~~
49 ~~since the management of the computing components is outsourced to IT experts. Or perhaps this is a narrative that~~
50 ~~the technology industry would like us to believe? This project is particularly interested in seeing if there has been~~
51 ~~changes over time in users’ experience of their ability to maintain digital information now that technologies like~~
52 ~~cloud computing are used widely.~~

53 **Methods**

54
55 This study will use employed a qualitative content analysis on $N=136$ participant responses to uncover the extent to
56 which there are changes in their experience of data loss using digital technology over time. For example, if digital
57
58
59
60

1
2
3 technology has indeed matured over time and services like cloud computing are helping preserve digital
4 information, we would expect to see a decline in responses that discuss extensive or catastrophic data loss.
5

6 These $N=136$ responses were collected from graduate students during the course of a graduate course in Library and
7 Information Science from 2013 to 2022. In this class, students responded to a discussion board post on personal
8 digital archiving and were asked in an open-ended manner, "Have you had any digital preservation challenges in
9 your past experiences (e.g., unable to access digital content) and what were they?" Students are encouraged to write
10 1-2 paragraphs each week on the discussion board. In some years, I also asked questions about readings related to
11 personal digital archiving, so in some cases students choose to discuss the readings rather than their personal
12 experience. This study will only be looking at responses related to the personal experience students had with digital
13 preservation, and will set aside responses that are outside of personal experience.
14

15 These discussion posts were collected from two learning management systems (Moodle and later Canvas), and
16 placed into a single text file. Each individual message was then added to a Google sheet as a row. To maintain the
17 privacy of respondents, the names were deleted from the dataset after some basic participant metadata was added,
18 such as year and semester the class was taken, gender pronouns and year of birth. Then, two graduate students were
19 selected to code the data. One coder was a graduate student in Library and Information Science with interests in the
20 area of archival studies. The second coder was a graduate student in a Master's program in user experience design
21 without any background knowledge on archives. The researcher discussed the codebook with the coders and asked
22 them to code each message for the experience of data loss reflected in the message, from no data loss to extensive
23 loss of information. They were also asked to code for why the respondent lost information, such as because of theft,
24 lost hardware, ~~lost hardware~~, failed hardware, service discontinued, damaged/destroyed hardware, among other
25 reasons for loss, including a combination of factors. The first coder made minor adjustments to the code book
26 through discussions with the researcher, and the second coder used this refined codebook when doing the coding.
27

28 Once the data ~~is was~~ coded, statistical analysis ~~will take place~~ was completed. This includes a linear regression to
29 determine if the passing of time is related to the severity of data loss, both using the year the course was taken, as
30 well as the year of birth. For example, if extensive data loss was on the decline, we would expect to see a linear
31 relationship with one or both of our time variables. Descriptive statistics will also be produced to shed light on this
32 relationship, and quotes from respondents will be provided to tie the statistics to individuals' own narratives.
33

34 Before getting to the results, we should note that Cohen's Kappa was run to determine if there was agreement
35 between the two coders, also known as inter-rater reliability. There was fair agreement between the two coder's
36 judgments, $\kappa = .397$. $p < .001$. For analyzing statistics from the coded discussion posts, ~~we will use~~ coder's number
37 one's coding ~~is used as, knowing that~~ it aligns fairly with coder number two's judgements. Using this coding, 56
38 responses were excluded from the dataset because they did not discuss personal experience, thus yielding $N=136$
39 responses.
40

41 In terms of demographics of the respondents, the average age of the students is between 29 and 30 years old. 78.4%
42 use she/her/hers pronouns, 17.9% use he/him/his pronouns, and 3.7% use they/them/their pronouns.
43

44 Results

45 Results from the qualitative content analysis indicate that there is no pattern over time related to personal data loss.
46 For example, if we look at the year the course took place and the average experience of personal data loss, with a
47 scale of "1" being no loss and "3" being extensive loss, we see no clear pattern, with the average going up and down
48 from 2013 to 2022, as shown in both Figure 1 and Table 1. Further, a linear regression confirmed that there is no
49 linear relationship between time and experience of data loss, both for the year the course was taken as well as the
50 year the student was born. Thus, the experience of data loss as time goes by is not on the decline.
51
52
53
54
55
56
57
58
59
60

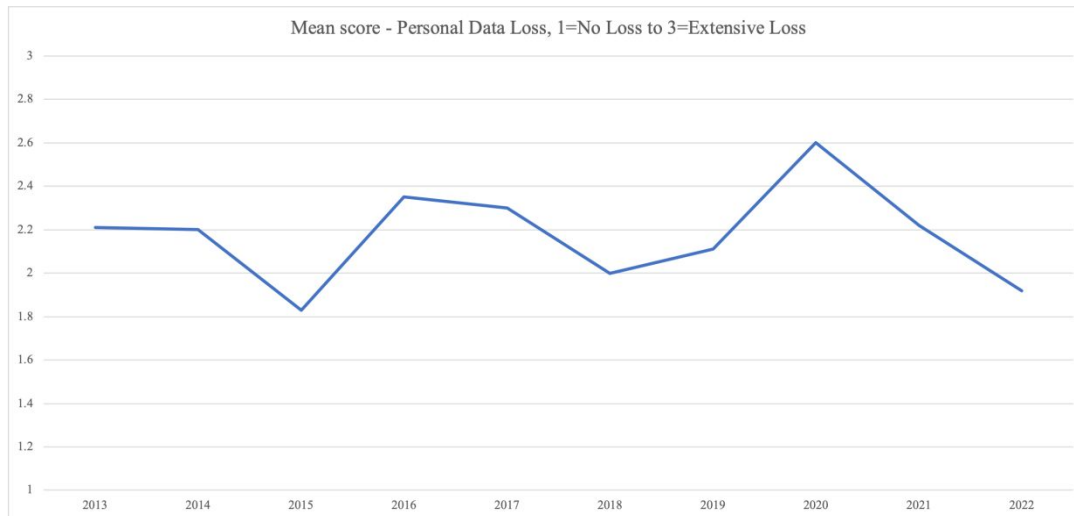


Figure 1. Mean score of personal data loss, which goes up and down over time, indicating no clear pattern

Table 1. Mean score of personal data loss with N respondents and standard deviations.

Year of Class	Mean - Personal Data Loss, 1=No Loss to 3=Extensive Loss	N	Std. Dev.
2013	2.21	28	0.738
2014	2.2	10	0.422
2015	1.83	6	0.753
2016	2.35	17	0.786
2017	2.3	10	0.823
2018	2	6	1.095
2019	2.11	9	0.601
2020	2.6	15	0.632
2021	2.22	23	0.6
2022	1.92	12	0.515

The qualitative narratives provide texture to the statistics, and shed light on the emotionally fraught nature of some of the experiences. For example, a 30-year-old student with they/them/their pronouns writes about a failed laptop in 2020:

I was gifted a laptop before undergrad, and I used it all the time. It had every essay, artwork and important document[s] stored on it's hard drive. The laptop functioned without a problem for 6 years. During that time, I never thought to back-up any files, nor did I purchase an external hard drive for the information. Eventually, the laptop began over-heating. I decided to invest in a new machine, and left the old laptop hoping that I would "get to it" another time. When I finally decided to work on transferring files and data to an external hard drive, the old laptop would not even start. It was really disappointing. I lost most of my reference photos for paintings, which I had decided later were actually more important than the paintings themselves. Now, I use two external terabyte hard drives. So many of my friends who are artists have one back-up external hard drive, but I have heard terrible anecdotes [sik] about theft, loss and functional deterioration [sik]. I choose to have two just for that reason. Storage seems to be an important part of personal digital archiving.

The above example highlights the negative emotions attached to these experiences of data loss (e.g., “really disappointing”). It also highlights that hardware failure continues to be an issue. While straightforward failed hardware is regularly experienced, often in the form of a failed laptop, the profusion of information across multiple devices and services can also open opportunities for data loss. For example, a 28-year-old respondent with she/her/hers pronouns discusses both loss on a mobile device and on social media in 2020:

My iPhone is always at least two generations out of date, which can cause problems with compatibility, and the asymmetrical storage space makes it difficult to keep my music on my phone. Also, like [name removed], I have had issues with not being able to move my music library from one computer to another! It can take a long time to amass and curate a library like that, and losing all that data is disheartening. / I also encountered a situation similar to [name removed]; I thought that my Facebook messages would be stored forever, since, for a period of time, I could go all the way back to the beginning of all of my conversations, and they were fully searchable. However, about a year or so ago, that stopped being the case. I don't know for sure if they're deleted after a set period of time or if the search function only goes back so far.

The above example highlights the negative emotions associated with losing data or being unable to access information (e.g., “losing all that data is disheartening”).

Services provided by commercial entities like Facebook are not the only source of data loss. Several students also report losing access to college email accounts upon graduating as a source of lost information. For example, a 28-year-old student with she/her/hers pronouns writes the following in 2020:

Oof, the experiences shared by my peers are painfully familiar. Like [names removed], I was somewhat blasé in navigating the transition from my undergraduate time. Despite receiving numerous recommendations from the institution to migrate my emails before my account was closed, I allowed my undergraduate email to disappear into the ether. I had not recognized how many of my professional and academic contacts were associated only with that account until I no longer had access to it. Since, I have been more mindful about which account I use for conversations and also have plans to migrate my [institution name removed] account information prior to graduation.

In the above example, recognition of a “painful” familiarity summarizes a collective experience of losing important information.

Of course, not all narratives included a story of catastrophic data loss. However, they point to other issues such as being overwhelmed with a profusion of information across devices and platforms. For example, the following was offered by a 30-year-old student in 2019 with she/her/hers pronouns:

...currently, I am dealing with the issue of having my digital assets being distributed over many different devices, some overlapping, and some completely independent. I obtained a new laptop in the summer of 2017, unlike some people I did not migrate all of the contents from my old laptop onto the new device, as I thought that I would transfer them onto a hard drive. On the old laptop contains coursework from my undergrad, other important documents, photos, etc. My new laptop has similarly important documents and photos, and some overlap with my old laptop. I also have an external hard drive that is over 10 years old and use cloud software, which holds materials from various moments in time. Copies that are distributed across different devices are the same, where the resolution, file format, and name remain the same. / However, with all my digital assets occupying different platforms, I run into the issue of not knowing where specific items reside. The challenge is getting to the material efficiently without having to open 4 different devices/software.

Thus, while some participants may have avoided a terrible data loss, there may be some anxiety around the proliferation of information across multiple devices.

In summary and with respect to research question #1, or has the occurrence of personal catastrophic data loss improved with the maturity of digital technology and services, we can see no evidence in this dataset that this is the case. With respect to the research question #2, or have the reasons for data loss changed with the maturing of digital technology, here we also don't see a clear pattern over time for the reasons for data loss, as indicated in Table 2. For

example, if we were to assume that individuals were using cloud computing to maintain their personal information, then we might expect to see a decline in issues related to hardware, or more issues related to “services.” This, however, is not the case and we still see “failed hardware” as being an issue reported frequently in both 2021 and 2022. This is also borne out in the qualitative narratives included above. Thus, a variety of factors, and in some cases a combination of factors, leads to personal data loss without any clear changes or trends over time.

Table 2. Reasons for data loss

Reason for Loss / Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
No loss or not discussed	5	0	2	3	2	3	1	0	2	2	20
Unspecified Loss	3	1	0	0	1	1	0	1	0	1	8
Lost Hardware	0	1	0	0	0	0	0	1	0	0	2
Failed Hardware	8	1	1	1	2	0	1	2	7	6	29
Service discontinued	2	3	0	1	0	1	0	0	2	2	11
Damaged/destroyed hardware	0	0	1	0	1	0	1	1	1	0	5
Data/information overwhelm	0	0	0	0	0	0	1	1	1	0	3
Outdated software	1	0	0	0	0	0	0	0	0	1	2
Outdated hardware	1	1	0	0	0	0	1	0	1	0	4
Improper Labeling	0	0	0	0	0	0	0	1	1	0	2
Combination of factors	8	3	2	11	3	1	4	8	8	0	48
Total	28	10	6	16	9	6	9	15	23	12	134

Discussion

This study found that as time has passed (from 2013-2022) that there has not been a decline in individual’s experience of data loss. While changes have occurred in the data storage area, such as the widespread use of cloud computing and solid-state storage, this has not led to improvement in the personal experience of data loss. Narratives around data loss still include negative emotions, such as pain and disappointment, indicating that catastrophic data loss continues to persist. The reasons for data loss also continue to vary without any clear trends, such as hardware failure, discontinued service, or a combination of factors.

If the technology itself is indeed maturing, why are we not seeing a decline in the experience of personal data loss? One potential reason is that while the technology itself might be getting better, individuals are running more parts of their lived experiences through digital technology, and for that reason there is simply more information that can be lost, forgotten or mismanaged. Whereas the literary figures Kirshenbaum studies like John Updike and Ralph Ellison may have had some small facet of their life on the computer (e.g., their literary production measured in a few kilobytes per week), today’s users are using digital technology to mediate vast segments of their lives. Even with better technology, data loss is still likely to occur because there is so much of it.

One limitation of this study is that the respondents are from a fairly limited demographic: graduate students (above average education), 29-30 years old (younger adults), and 78.4% using she/her/hers pronouns. For this reason, caution should be used in applying these findings to larger and more diverse populations. A further limitation of this study is that it only looked at changes over a time window of nine years (2013-2022), and it may be worthwhile to revisit these experiences of personal data loss over a longer time frame, such as 15 or 20 years.

Conclusion

In conclusion, this study found that despite the advances in digital technology, personal computing users over time continue to experience catastrophic personal data loss. While the technology itself may be advancing, because more facets of life are mediated through it, there are simply more opportunities for information to be lost, and thus there is no downward trend in data loss. The reasons for data loss continue to be varied, such as hardware failure and discontinued services. Thus, the main implication of this research is that any computing user who wants to avoid the negative emotions associated with personal data loss needs to enact or continue to enact personal digital archiving practices, and the fact that personal and mobile computing have been with us for some time is not reason to let one’s guard down.

1
2
3 While all of the study subjects were college graduates, none of them mentioned learning about personal digital
4 archiving in school, either in secondary school or college. This would indicate that there may be opportunities to
5 expose students to personal digital archiving at a younger age, perhaps through information literacy or computing
6 skills curriculum. Many students in the study learned about data loss “the hard way”—through actually losing their
7 data—and this need not be the case. As this study found that these experiences of data loss are accompanied by
8 negative emotions, such as pain and disappointment, and that these records may be useful in recounting one’s life as
9 a digital “memory closet,” then exposing students to basic curriculum on personal digital archiving may be
10 worthwhile, such as inventorying where important records are located, and if there are multiple copies. Basic
11 education on personal digital archiving can help alleviate “that sting of avoidable loss” and buttress young peoples’
12 future digital memory closets.

13
14 While this project focused on personal digital archives, some implications can be drawn for collecting institutions.
15 It behooves collecting institutions to work with donors on their personal digital archiving practices or else risk a data
16 loss incident before the material is ever transferred. For example, helping donors develop an inventory of the
17 location of their personal digital collections (e.g., old computers, cloud services, etc.) so that the risk to them can be
18 evaluated would be a worthwhile first step. Important materials may be on aging hardware or failing cloud services,
19 jeopardizing their longevity and the ability to ever donate them. Further, developing public programming and
20 messaging to promote personal digital archiving practices as a general community service is also worthwhile as
21 community members become prospective donors based on ever-evolving collecting policies.
22
23
24

25 References

- 26
27 Becker, Devin and Collier Nogues. “Saving-Over, Over-Saving, and the Future Mess of Writers' Digital Archives: A
28 Survey Report on the Personal Digital Archiving Practices of Emerging Writers.” *American Archivist* 75.2
29 (Fall/Winter 2012): 482-513.
30
31 Bureau of Economic Analysis. “What is considered a catastrophic loss?” <https://www.bea.gov/help/faq/1354>
32 (accessed 3 June 2022).
33
34 Coccio, Anthony. “Youth Deleted: Saving Young People’s Histories after Social Media Collapse.” Paper
35 presented at the *International Internet Preservation Consortium General Assembly*, May 19-23, 2014,
36 Paris, France. http://www.thinkingprojects.org/youth_deleted_iipc.pdf (accessed 7/1/2022).
37
38 Gunn, Chelsea. “Introduction: Putting Personal Digital Archives in Context.” In *The Complete Guide to Personal*
39 *Digital Archiving*, edited by Brianna H. Marshall, xi-xxii. Chicago: ALA Editions, 2018.
40
41 Kirschenbaum, Matthew G. *Track Changes: A Literary History of Word Processing*. Cambridge, MA: Belknap
42 Press of Harvard University, 2016.
43
44 Lee, Christopher A. I. *Digital: Personal Collections in the Digital Era*. Chicago: Society of American Archivists,
45 2011.
46
47 Marshall, Brianna H. (Ed). *The Complete Guide to Personal Digital Archiving*. Chicago: ALA Editions, 2018.
48
49 Marshall, Catherine C. “Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field.” *D-Lib*
50 *Magazine* 14.3/4 (March/April 2008). <https://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html>
51 (accessed 7/1/2022).
52
53 Millar, Laura. “Touchstones: Considering the Relationship between Memory and Archives.” *Archivaria* 61
54 (September 2006): 105-126.
55
56 Post, Colin. “Preservation practices of new media artists: Challenges, strategies, and attitudes in the personal
57 management of artworks.” *Journal of Documentation* 73.4 (2017): 716-732.
58
59
60

1
2
3
4 Turkle, Sherry. *The Empathy Diaries: A Memoir*. New York: Penguin.
5

6 “Data loss.” Wikipedia article, available at https://en.wikipedia.org/wiki/Data_loss (accessed 6/3/2022).
7

8 Zhao, yue, Xian’e Duan and Haijuan Yang. “Postgraduates' personal digital archiving practices in China: Problems
9 and strategies.” *Journal of Academic Librarianship* 45.5 (2019): 1-11.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 September 24, 2022
4

5
6 Dear Dr. Gracy,
7

8 Thank you for taking the time to have this article peer-reviewed. Please find with this letter the
9 revised manuscript. Also, you can find how the manuscript below was updated to respond to
10 the reviewer comments:
11

12
13 Review 1

- 14 - *"lost hardware' are repeated"* – The repetition has been corrected.
- 15 - *"Methods seems to be written in the wrong tense"* – The tense has been standardized to
16 the past tense.
- 17 - *"I would suggest that the conclusions/discussion section could offer some suggestions
18 for collecting archives/memory institutions regarding how the findings might influence
19 their communications with personal donors/depositors in terms of influencing their
20 personal digital archiving attitudes and practices."* – A paragraph was added to the
21 conclusion to address the implications for collecting institutions.
22
23
24

25
26 Review 2

- 27 - *"The maturing of digital technology"* – The discussion of the maturing of digital
28 technology has been moved out of literature review and into the introduction.
- 29 - *"Tenses should be standardized throughout"* – The tenses have been standardized to the
30 past tense.
- 31 - *"It would seem that the data collected from the student responses does not readily lend
32 itself to statistical analysis, or to analyzing any correlation between year/date and
33 severity of data loss. This component of the analysis is confusing since year/date is not a
34 quantitative variable, and the scores associated with personal data loss represent
35 categorical data rather than data which shows any quantitative distinctions."* – I have
36 re-reviewed the statistical analysis presented in the paper and consulted some sources
37 and I do stand behind its use. Both the year and the severity of data loss are
38 quantitative variables despite the reviewer's suggestion that they are categorical
39 variables. Quantitative data is data that can be measured and those differences are
40 meaningful. For example, the year 2020 is five years after the year 2015, indicating a
41 passage of time of five years. Using years as a quantitative variable is frequent in
42 statistics, such as other year-based quantitative variables like age (the number of years
43 someone has been alive). The severity of data loss has also been coded into a
44 quantitative variable ranging from 1 to 3, where "1" indicates "no loss" and "3"
45 indicates "extensive loss." This is also a quantitative variable, where "3" is more loss
46 than "2", "2" is more loss than "1", etc. There is categorical data used in this study—the
47 reasons for data loss—however, that data is not part of the statistical analysis. The
48 chart in Figure 1 illustrates the quantitative nature of this data, and also shows there is
49 no linear relationship between time-passed and the severity of data loss.
50
51 - *"The reasons for data loss could be expanded upon to include some discussion and
52 analysis of each of these codes, and what kinds of experiences are represented within
53
54
55
56
57
58
59
60*

1
2
3 *each of these codes/categories. It would also be useful to know how codes were*
4 *assigned. Were they predetermined, or was open coding.”*

- 5
6 - *“The implication of the findings, who the research might benefit in future, and how the*
7 *study contributes to the realm of personal digital archiving is not clear to the reader” - I*
8 *have added a sentence to the conclusion that makes it clear that benefit of the research*
9 *is to let people know they need to practice (or continue to practice) PDA because*
10 *personal data loss is not on the decline. Also, as requested by the first reviewer, I have*
11 *added implications for collecting institutions, which should help make clear the benefits*
12 *of the research.*

13
14
15 Thank you again for your work on this article.

16
17 All the best,

18
19
20
21 Anthony Cocciolo
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60