Digital hoarding and personal use digital data

Elizabeth Sillence[1,2]*, Jordan A. Dawson[1], Richard D. Brown[1], Kerry McKellar[1,2], Nick Neave[1,2]

[1]Psychology Department, Northumbria University, Newcastle upon Tyne NE1 8ST

[2]North East Hoarding Research Group, Psychology Department, Northumbria University, Newcastle upon Tyne NE1 8ST

*Elizabeth.sillence@northumbria.ac.uk

Abstract

The problem of 'Digital Hoarding' has been studied in a workplace context but less so in relation to personal use digital data. In this study, we examine the relationship between scores on a measure of Digital Hoarding and the digital behaviors associated with personal technology. 167 participants completed the Digital Hoarding Questionnaire (DHQ) to assess digital hoarding tendencies. Participants indicated their main storage device for personal use digital data and thought about this when completing a new measure – the Digital Behaviors for Personal use Questionnaire (DBPUQ) designed to capture information item management behaviors and affective responses to digital information items. Higher scores on the DHQ correlated with difficulty finding information items, attachment and distress at potential loss of information items. Participants' choice of personal device was not associated with DHQ scores, information item management behaviors, affective responses or total information items stored. This work provides a first step towards a better understanding of personal use digital data hoarding. We discuss the implications of digital hoarding of personal use data with respect to increased feelings of anxiety, cybersecurity vulnerabilities, and sustainability concerns. Understanding digital hoarding in a personal use data context is foundational for developing design solutions and behavioral interventions aimed at reducing digital data overload.

Keywords, digital hoarding; personal information management, smartphone use; personal use digital data; deletion activity

1. Introduction

The tendency to accumulate digital information items and a failure to delete unnecessary items leads to digital clutter. Digital clutter can have negative impacts on productivity, as individuals struggle to locate files and applications, creating feelings of anxiety (Sweeten, Sillence, Neave, 2018., Sedera & Lokuge, 2018) and posing significant risks for both individuals and organizations in terms of potential data breaches. The extent to which individuals engage in 'digital data hoarding' varies. In workplace settings, job demands and organizational culture can impact on the volume of accumulated data (McKellar, Sillence, Neave & Briggs, 2020; 2023). Digital hoarding tendencies also differ between individuals. Vitale, Janzen & McGrenere (2018), for example, describe how people's behaviors in relation to their digital data can be thought of on a continuum between minimalism and hoarding.

Few studies have examined digital hoarding behaviors in relation to personal use digital data, as opposed to data used for work purposes. One exception is a study by Sweeten et al., (2018) who noted individuals' emotional attachment to personal information items, in particular photos and music. Participants in that study expressed a sense of anxiety about the thought of losing or accidentally deleting those personal information items (Sweeten et al., 2018). This qualitative study highlighted both the motivations and potential outcomes of digital hoarding but did not explore in detail the scale of digital hoarding in a personal use context. We believe that understanding digital hoarding in a personal use data context, and the specific behaviors and affective responses around that data, is a necessary starting point for developing products and behavior-based interventions to reduce digital data overload. This will help individuals reduce the feelings of anxiety linked to large volumes of unnecessary data and encourage more consideration of the sustainability issues associated with large volumes of digital data.

In this paper we introduce the use of a questionnaire designed to measure patterns of digital hoarding behaviors associated with personal technology (the Digital Behaviors for Personal Use Questionnaire: DBPUQ). Findings from this questionnaire provide an understanding of the scale of digital hoarding, specific behaviors relating to deletion and organization, and the affective responses to personal digital data. Furthermore, we use a validated measure of digital hoarding tendencies (DHQ) to examine the expected positive relationship between general digital hoarding tendencies and digital behaviors associated with personal technology. Investigating this expected relationship may help to validate the potential use of the DHQ for predicting hoarding behaviors associated with personal use digital data. This work represents an initial step in enhancing our understanding of personal use digital data hoarding, and how to predict, measure, and address these digital behaviors.

Therefore, the overall aims of the study are: (1) to examine the relationship between digital hoarding tendencies and digital behaviors associated with personal technology; (2) to examine whether digital behaviors associated with personal technology vary in relation to scores on the Digital Hoarding Questionnaire (DHQ); and (3) to examine how people's choice of device for storing personal digital data impacts digital data hoarding.

These aims are explored through three research questions:

RQ1: Do scores on the Digital Hoarding Questionnaire (DHQ) correlate with information item management behaviors, affective responses to data, and number of information items stored in relation to participants' chosen personal technology device?

RQ2: Do high and low scorers on the DHQ differ in their information item management behaviors, affective responses to data and number of items stored on a chosen personal technology device?

RQ3: Does choice of personal technology device affect information item management behaviors, affective responses to data and number of items stored?

We begin by reviewing the literature on digital hoarding, which has predominantly focused on workplace contexts. We then review the literature around device use in relation to Personal Information Management (PIM) and consider its implications for digital hoarding.

2. Background

2.1 Overview of digital hoarding literature

Initially described through a single case study, a small but growing body of research is now seeking to understand more about digital hoarding, (Sweeten et al., 2018, McKellar et al., 2020). Some researchers have focused on the theoretical foundations of digital hoarding (Sedera et al., 2022), while others have examined the underlying motivations for hoarding behaviors including, emotional attachment, keeping information 'as evidence' and time pressures (Sweeten et al., 2018). Other papers have focused on how data hoarding practices vary within specific contexts (McKellar, et al., 2023; Sillence, Dawson, McKellar & Neave, 2022). Researchers have also identified a number of possible technical factors that might make data hoarding easier, for example easy copy and paste functions and abundant storage (Dinneen, Julien & Frissen, 2019; Sweeten et al., 2018). Finally, a number of studies have noted the negative consequences of digital hoarding, particularly in relation to stress and anxiety (Sedera et al., 2022; Sweeten et al., 2018; Sillence et al., 2022).

2.2 Measuring digital data hoarding behaviors

Alongside ways of capturing digital hoarding tendencies and related behaviors, some researchers have developed file-specific scales (e.g., the digital photo hoarding scale (Bozacı, & Gökdeniza, 2020)) while others measure the overall tendency towards Digital Hoarding (Sedera, Lokuge & Grover, 2022, Neave, Briggs, McKellar & Sillence, 2019). The Digital Hoarding Questionnaire (DHQ; Neave et al., 2019), for example, measures digital hoarding tendencies and is based on the physical hoarding literature that sees two factors underpinning the issue – one around difficulty deleting and the other around accumulating. The DHQ has been shown to predict workplace digital data behaviors, such that employees scoring more highly on the DHQ had more stored files, particularly emails, and were less likely to delete information items (Neave et al., 2019).

2.3 Affective responses to digital data

The affective experience of managing one's information is also beginning to be recognized more widely (Alon & Nachmias, 2020) and is an important factor in decision making around information item deletion. We know that some information items may be better organized than others, and some items are more likely to be kept than others. For example, digital information items vary in their emotional resonance for individuals, and photos and personal creative writing documents are often highlighted as important digital possessions to preserve (Cushing, 2011). Sweeten et al. (2018) noted that people find it difficult to delete music files and photos, and we know that storing personal data is important in reminiscence or legacy creation (Thomas & Briggs, 2016; Gulotta, Odom, Forlizzi, & Faste, 2013).

Data collections in workplace or academic settings can also generate strong feelings. Kaye et al. (2006) noted that physical data collections build legacy and preserve information, while digital collections also develop and strengthen identity and purpose (McKellar et al., 2020). Students can exhibit a strong sense of attachment to their digital data, particularly with respect to assignments or projects they created themselves (Sillence, Dawson, McKellar & Neave, 2022).

2.4 The different contexts of digital data management research

Given the large volumes of data involved, and the cybersecurity risks associated with poorly managed digital data (Gormley & Gormley, 2012; Neave et al., 2019), it is unsurprising that the predominant focus of research has been on examining digital information collections through a workplace or educational lens, especially in relation to digital hoarding (Oravec, 2018; McKellar et al., 2020; Alon, Forkosh-Baruch, & Nachmias, 2020). Some researchers, for example Vitale et al., 2018, have focused on digital data management across a broad range of settings, including personal data, but less is known about personal use digital data and the associated digital hoarding behaviors. A number of studies by Dinneen and colleagues provide some insight into the types of information items contained within personal use collections. For example, Dinneen & Julien (2019) were the first to examine purely 'personal collections' – collections used for personal matters as opposed from those used for work and study. They noted that in comparison to work based or educational collections, personal collections contained more image and audio files, and that text-based files and spreadsheets were almost non-existent. In a separate study focusing on email, Dinneen & Krtalić (2020) discuss how emails can also form part of personal collections, and decisions around what to keep and delete can form part of how people come to tell their life story.

2.5 The role of device in digital data hoarding

While researchers are beginning to understand more about the relevance of certain information items, in particular email, there has been little consideration of device in the digital hoarding literature. In a work context this is understandable given that most employees use a laptop or desktop computer. However, in relation to personal data the range of devices individuals may choose to use is likely to be broader and include smartphones. In this paper we use the term 'personal device' to refer to devices such as smartphones, laptops or PCs that are personally owned rather than work based. The increasing availability of personal devices is likely to exacerbate the volume of stored digital data (Sedera et al., 2022) and PIM researchers provide interesting insights into device-based differences in relation to file retrieval. Bergman & Yanai (2018), for example, found that retrieval performance for emails and files was worse for smartphones compared to computers. However, photos may be different. Bergman, Gutman & Whittaker (2022) looked at how smartphones affected picture retrieval, assessing two conflicting factors – the increased size of picture collections on smartphones and the potential for emerging technologies to aid picture retrieval. They noted that collection size on smartphones increased retrieval failures compared to those in an earlier study utilizing digital cameras. However, participants were able to make use of timeline features, search and face recognition to locate photos on smartphones.

Smartphones are increasingly relevant for PIM research as younger generations in particular take advantage of their storage capacity and portability to underpin their personal information management behaviors (Zhang & Liu, 2015). Studies of personal information management using smartphones suggest a lack of systematic methods for keeping and organizing information (Ali & Warraich, 2021). Smartphones may also be used as part of a wider approach to PIM with students engaging in what Ali & Warraich (2022) refer to as 'sharing as keeping', where information is sent via SMS, WhatsApp and email to themselves and others as a way of keeping the information. Creegan

(2017) and Sillence et al. (2022) both note the relatively common practice of screenshotting information and emailing self and others. These practices lead to the duplication of digital data, as highlighted by Henderson (2011). Furthermore, research on smartphone usage within PIM has important insights for digital hoarding. It suggests that the increase in data creation, greater storage capacity, and difficulty retrieving information items all add to a potential burden. We anticipate that device type impacts hoarding behaviors and affective responses to data. We also expect that smartphones are associated with greater hoarding behaviors and stronger affective responses to the data stored on these devices.

In summary, research into digital hoarding has mainly focused on work-based or educational contexts. The literature has paid limited attention to the understanding of digital hoarding tendencies, behaviors and affective responses associated with personal use digital data. Little is known about digital hoarding tendencies associated with personal use digital data, and the impact of people's choice of device type on this form of digital data management. Understanding more about the scale of digital hoarding in a personal use context is important foundational work if interventions to reduce digital data overload are to be achieved. Understanding the role of device type in digital data hoarding is also key to the development of appropriate solutions to reduce unnecessary data storage that are tailored to individual devices.

3. Method

3.1 Preliminary study

Before developing the materials for the main study, we conducted a small-scale preliminary focus group (n=5) to understand more about personal use digital data and personal devices. The aim of this focus group was to see what users understood by the term 'personal device', to explore information item types relevant to personal digital data and to explore issues surrounding cloud storage. Participants were invited to discuss their use of personal devices in the home, their digital information items, and their associated hoarding behaviors. This focus group helped the research team to identify the types of digital information items to present in the DBPUQ, as well as the specific wording around the instructions concerning choice of personal device. Two main findings from the focus group concerned devices and cloud storage. Firstly, it was clear that some devices were frequently used but did not necessarily contain the majority of an individual's information items. In order to keep our focus on information items, we decided to ask participants to choose a device that contained the majority of their information items. Second, there was large variation in what was understood by the term 'cloud storage' and how people operationalized any use of cloud storage including differences in synchronization and automatic backups. Furthermore, many people only considered cloud services in relation to their work data and talked about their personal devices when describing their use of personal digital data. We therefore decided to ask participants to focus on a specific device, as described above, and not consider cloud storage.

3.2 Main study

Sample

202 participants took part in the study, and 167 completed the questionnaire in full. Of those 167, 26 participants identified as male (aged 18 – 63, Mean = 31.35, S.D. = 15.73), 135 participants identified as female (aged 18 – 71, Mean = 26.81, S.D. = 11.10) and 6 participants identified as a

third gender or non-binary (aged 19 – 41, Mean = 24.33, S.D. = 8.34). The overall mean age was 27.43 (range = 18 – 71, S.D. = 11.91). Participants were recruited nationally and internationally via social media using the research team's individual and lab based social media accounts and also via the University's student recruitment portal.

Materials

The Digital Hoarding Questionnaire (DHQ): The DHQ (Neave et al., 2019) (see Appendix) is a 10-item questionnaire that measures two key factors associated with digital hoarding, 'accumulation' and 'difficulty deleting'. Six items measure difficulty deleting and four items measure accumulation. Example items include "I lose track of how many digital files I possess" and "Deleting certain files would be like losing part of myself". Responses are provided on a seven-point Likert Scale ranging from 1 (not at all) to 7 (very much so). Both sections of this questionnaire were found to have high internal consistency (Neave et al., 2019).

The Digital Behaviors for Personal use Questionnaire (or DBPUQ) (see Appendix) was adapted from The Digital Behaviors in the Workplace Questionnaire (DBWQ) (Neave et al., 2019). The DBWQ was created to assess the extent of digital hoarding in the workplace. We adapted this questionnaire (following the findings of the preliminary study) by extending the types of information items to include apps, eBooks, movies and music, while removing presentations. We also removed the section on implications of hoarding behaviors for the individual and the work organization. The newly adapted questionnaire captures hoarding behaviors, and data on the number and type of information items stored and focuses on personal devices and personal information items. Participants were asked to choose a personal device that contained the majority of their information items (this might not be the device they use most often), and to think about that device when considering the items they stored, their deletion and information item management behaviors and their affective responses towards their data. The DBPUQ captured the following data:

> *Digital information item types*. We asked participants to state how many of the following information item types were stored on their chosen device including: photos, music, apps, eBooks, read emails, unread emails, text-based files and numerical files e.g., excel spreadsheets.
>
> *Deletion activity.* Participants were asked how often they delete each type of information item (see above for the digital information item types) on a five-point scale of daily (1), weekly (2), monthly (3), yearly (4) to I hardly ever delete these files (5). Higher scores indicate participants are reluctant to delete and therefore keep more digital information items. With each box allocated a numerical value of 1–5 (daily to hardly ever) then a total summed score was calculated for deleting behaviors, a higher score indicating greater hoarding behavior (less deleting activity).
>
> *Information item management behaviors.* Participants were asked two questions addressing organizing and retrieval behaviors with respect to the information items they have stored on their chosen personal device. Responses were provided on seven-point Likert scales. The first question asked, "To what extent do you organize these items?" (1 = no organization at all to 7 = very well organized). The second question asked, "To what extent do you have difficulty finding these items again?" (1 = no difficulty at all to 7 = very difficult indeed). A total summed score across the information item types was calculated to give a single score for 'organization behaviors' and 'difficulty finding'.

Affective responses to digital information items. Participants were asked, "For each information item type tell us how strongly attached to them you are, and how distressed you would feel if they were deleted/lost forever." Participants responded on seven-point Likert scales from 1 = not at all to 7 = very much so. A total summed score across the file types was calculated to give a single score for 'attachment' and 'distress'.

Procedure

Following approval from our institutional ethics committee [reference number 17207], participants completed an online survey consisting of demographic information, the DHQ and the DBPUQ. The survey took approximately 15 minutes to complete, and participants did not receive any financial renumeration for taking part in the study.

Analysis

Statistical analyses were conducted using SPSS and R (IBM Corp, 2020; R Core, 2021). The psychometric properties of the DHQ (Neave et al., 2019) were assessed by first conducting a confirmatory factor analysis to determine whether the previously suggested two-factor solution (Factor 1 = difficulty deleting, and Factor 2 = accumulating) provided optimal model fit for the data (see below for full results). A subsequent principal component analysis showed that one component accounted for a substantial degree of the variance in participant responses. Therefore, a single summed score for all DHQ responses was used as our measure of digital hoarding. Using a single summed score follows recommendations that summed scores are most acceptable when using newly developed scales and help to preserve the variation of the original data (DiStefano, Zhu & Mindrila, 2009).

For RQ1, Pearson correlations were conducted between total DHQ scores, deletion activity, information item management (organization and difficulty deleting), affective response (distress deleting and attachment), and total number of items stored on participants' chosen device. For the DHQ and DBPUQ variables, total scores were summed and treated as continuous variables following recommendations for treatment of multiple Likert responses in social science research (Subedi, 2016; DiStefano et al., 2009).

However, data for the number of information items held on participants' personal devices were not normally distributed. Therefore, for correlation tests including the total number of information items stored, the R package 'bestNormalize' was used to transform the non-normally distributed data (Peterson & Peterson, 2020). BestNormalize performs a suite of normalizing transformations, then selects the transformation that has the most appropriate test statistic for normality. Previous research has shown that, for most sample sizes ($n \geq 20$), Type I and Type II error rates are minimized by transforming the data to a normal shape prior to assessing the Pearson correlation (Bishara & Hittner, 2012).

For RQs 2 and 3, t-tests were conducted to investigate differences in responses for DBPUQ components, and total information items stored, between high and low hoarders and between those who did/did not select mobile phone as their preferred device. The normalized variable for the total number of information items stored was again used when conducting t-tests as previous research has found that transforming non-normal data before performing a t-test protects against changes in the probability of Type I and Type II errors that occur in the case of several long-tailed distributions (Zimmerman & Zumbo, 1993).

4. Results

We began by checking the psychometric properties of the DHQ in our sample. A Confirmatory Factor Analysis (CFA) was conducted to evaluate the previously defined two-factor structure (Factor 1 = difficulty deleting, and Factor 2 = accumulating). The factor loadings for all items on their respective factors were significant at $p < .001$, demonstrating that all items were significant indicators of their respective factors. However, the model fit was less than optimal, with CFI = 0.872, TLI = 0.831, RMSEA = 0.163, and SRMR = 0.079, suggesting the model did not perfectly fit the data. Despite this suboptimal model fit, values for Cronbach's alpha for both factors showed similar levels of high internal consistency: Factor 1: Difficulty deleting (6 items) $\alpha = .912$, Factor 2 accumulating (4 items) $\alpha = .864$ (Taber, 2018). A subsequent principal components analysis showed that a single component explained 61.12% of the variance in the 10-item DHQ, indicating that a significant proportion of the responses can be accounted for by a single component. Each subsequent component explained considerably less variance, with the second component accounting for only 11.8%. Furthermore, Cronbach's alpha for the full DHQ demonstrated an excellent level of internal consistency, $\alpha = 0.93$ (Taber, 2018). This supports the use of an overall DHQ score, taking the summed values of each of the 10 responses. Therefore, the 10 DHQ survey responses were summed to produce a single score for digital hoarding. A total summed digital hoarding score was also used in the original validation of the DHQ (Neave et al., 2019).

We then moved on to examine our three study research questions separately.

RQ1: Do scores on the Digital Hoarding Questionnaire (DHQ) correlate with information item management behaviors, affective responses to data, and number of information items stored in relation to participants' chosen personal technology device?

We examined whether the digital hoarding behaviors in general (as measured by the DHQ) predicted digital behaviors associated with personal technology (as measured by the DBPUQ). Overall, there was a significant positive correlation between total scores on the DHQ and a participants' overall deletion activity ($r = .384$, $p < .001$) in the DBPUQ.

In terms of information item management behaviors, participants' scores on the DHQ did not correlate significantly with a participant's tendency to organize their data in the DBPUQ ($r = -.056$, $p = .478$), although there was a significant correlation between DHQ scores and difficulty finding information items ($r = .483$, $p < .001$).

In terms of affective responses, scores on the DHQ also showed a significant positive correlation with attachment to information items in the DBPUQ ($r = .400$, $p < .001$) and distress if item lost ($r = .367$, $p < .001$).

Table 1. Means, standard deviations, and correlations between DHQ, DBPUQ, and total information items stored.

Variable	*Mean*	*S.D.*	1	2	3	4	5	6
1. DHQ Total	39.51	15.39						
2. Deleting Behaviors	37.47	9.88	.38** [.25, .51]					

3. Organize Behaviors	38.79	14.53	-.06 [-.21, .10]	.20** [.05, .35]				
4. Difficulty Finding	34.63	13.32	.48** [.35, .59]	.49** [.36, .60]	.09 [-.07, .24]			
5. Attachment Behaviors	86.83	22.88	.40** [.26, .52]	.56** [.44, .65]	.40** [.26, .52]	.54** [.42, .64]		
6. Distress	43.48	11.66	.38** [.24, .50]	.53** [.41, .63]	.37** [.23, .50]	.51** [.39, .62]	.97** [.95, .98]	
7. Total Items Stored	24,053	91,567	.38** [.24, .50]	.30** [.16, .43]	.09 [-.07, .24]	.23** [.08, .37]	.29** [.14, .42]	.25** [.10, .39]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$. The total files scored variable was normalized to satisfy the normality assumption of a Pearson correlation (see section 'analysis' above).

Table 2 indicates the mean and standard deviation of the number of information items stored on participants' chosen device for each of the different information item types. The table shows that some information item types, in particular emails and photos, are stored in large numbers on personal devices. It is also clear that the number of information items stored varies enormously between individuals. DHQ total scores were positively associated with the total number of items participants had stored on their personal device (data were normalized to satisfy test assumptions, see analysis section above; r = .381, p <.001).

Table 2. Mean and standard deviation of stored information item types

Information item Type	*M*	*SD*
Photos	6272	13215
Music	1502	7687
Apps	66	65
Videos	637	1318
Movies	17	156
e-Books	94	672
Read Emails in In-Box	9287	77800
Unread Emails in In-Box	1961	6332
Deleted Emails	274	898
Archived Emails	2958	19667
Text Files	906	4850
Numerical Files	94	490

RQ2: Do high and low scorers on the DHQ differ in their information item management behaviors, affective responses to data and number of information items on a chosen personal technology device?

We investigated differences between high and low scorers on the DHQ in terms of: total information items stored, deletion activity, information item management behaviors and affective responses as measured by the DBPUQ.

Participants were split into high and low digital hoarders based on their DHQ score. A participant was deemed a "digital hoarder" if they scored more than 16 on Factor 1, Difficulty Deleting, AND more than 15 on Factor 2, Accumulating (based on the approached used by McKellar et al., 2020).

A total of 97 participants were classified as high scorers, and the remaining 70 were classified as low scorers. 16 men (61%) scored above the cut-off point, 76 (56%) of women and 5 (83%) of non-binary participants.

Information items stored

There was a statistically significant difference between the two groups (high versus low DHQ scorers) in terms of the total number of information items (data were normalized to satisfy test assumptions), $t(149.55) = -4.448$, $p < .001$ ($Mean_{low\ DHQ} = 8{,}024$, $Mean_{high\ DHQ} = 35{,}622$). See Figure 1 for differences between low and high scorers on the DHQ for number of files stored by file type.

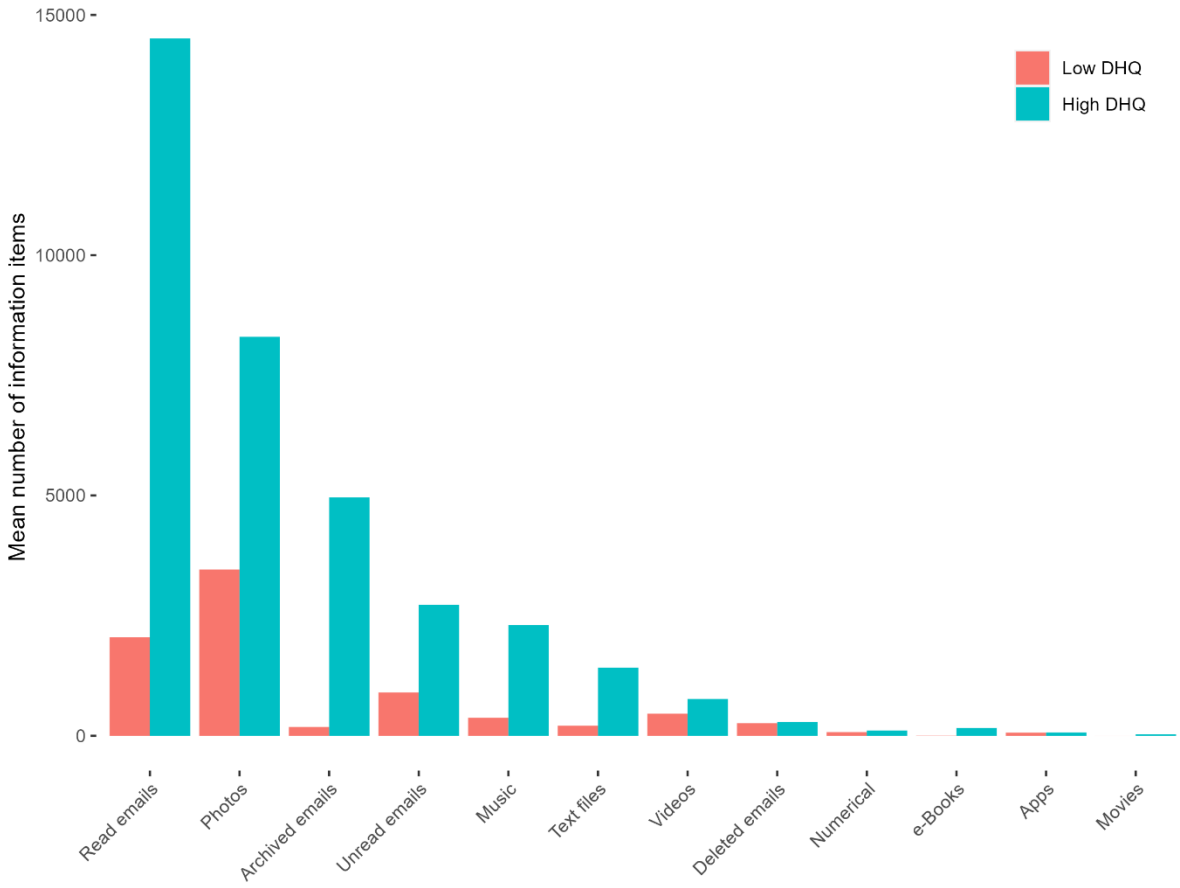

Figure 1. Number of information items stored by low and high scorers on the Digital Hoarding Questionnaire (DHQ).

Deletion activity, information item management behaviors and affective responses

There was no significant difference between high and low scorers on the DHQ regarding the extent to which they organized their information items overall. However, the other behaviors and affective responses varied significantly between the two groups (high and low scorers) as shown in Table 3.

Table 3. Differences in DBPUQ responses between high and low scorers on the DHQ.

DBPUQ measure	t-test result	Mean (High DHQ)	Mean (Low DHQ)
Deleting activity	t (165) = -4.198, p <.001	40.07	33.87
Attachment	t (165) = -4.200, p <.001	92.86	78.49
Difficulty Finding	t (160) = -5.024, p <.001	38.75	28.79
Distress	t (164) = -3.778, p <.001	46.28	39.59
Organization	t (164) = 1.155, p =.250	37.69	40.33

Note. High deleting score means participants are more reluctant to delete information items.

Overall, high scorers were less likely to delete information items, had more difficulty finding items, were more attached to information items and were more distressed at the thought of losing items.

RQ3: Does choice of personal technology device affect information item management behaviors, affective responses to data and number of information items stored?

To examine any differences in relation to the device chosen by participants we created two different device groups, (i) smartphone and (ii) other. In total, 106 people chose their smartphone and 59 people chose another device (2 participants did not complete) including Laptop (N=41), Hard drive (N=7) PC (N=4) iPad (N=4) USB drive (N=2) Nintendo switch (N=1).

Information items stored

There was no difference between the smartphone group and the 'other device' group in terms of the total number of information items stored, t(113.6) = 0.395, p = 0.694.

High and low scorers on the DHQ

There was no difference between high and low hoarders regarding the device they chose, χ^2 (1, 167) = .7, p = .403.

Deletion activity, information item management behaviors and affective responses

There were no differences between the device chosen and the deletion, information item management and affective responses scores (see Table 4).

Table 4. Differences in DBPUQ responses by chosen personal device.

DBPUQ measure	t-test result	Mean (Phone)	Mean (Other)
Deleting activity	t (163) = 0.620, p =.536	37.00	38.00
Attachment	t (163) = -0.956, p =.341	88.00	84.41
Difficulty Finding	t (158) = -0.065, p =.948	34.70	34.55
Distress	t (161) = -1.332, p =.185	44.39	41.84
Organization	t (161) = -0.330, p =.742	38.99	38.21

Summary of findings

This study has focused on digital data management in relation to personal technology. Specifically, in relation to the first study aim we found that scores on the DHQ, a measure of general digital hoarding, correlated with digital behaviors associated with personal technology as measured by the DBPUQ. Higher scores on the validated DHQ correlated with difficulty deleting, difficulty finding information items, attachment to information items and distress at potential loss of items. We also found a correlation between the total number of information items stored and higher scores on DHQ.

Looking at high and low scorers, we found that high scorers (digital hoarders) were less likely to delete information items, had more difficulty finding items, were more attached to items, and were more distressed at the thought of losing information items. There was also a significant difference between the two groups in relation to the overall number of items stored, with higher scorers storing more information items on their personal device.

Finally, in relation to the chosen personal device, we note that the majority of participants chose their smartphone. However, there were no differences in terms of device chosen and information item management behaviors, deletion activity and affective responses. Furthermore, there was no difference between high and low scorers on the DHQ and device chosen. There was also no difference between smartphone and other devices in terms of the total number of information items stored.

5. Discussion

We have successfully shown that scores on the DHQ correlate with digital behaviors as measured by the DBPUQ. This shows that the DHQ (a measure of digital hoarding tendency overall) may provide a useful predictor of digital hoarding behaviors associated with personal use digital data. The DBPUQ was able to differentiate between high and low scores on the DHQ and this makes it a potentially useful tool for assessing hoarding behaviors, affective responses and information item types across further personal use contexts, including in relation to family-based information management (Krtalić, Dinneen, Liew, & Goulding, 2021; Sannon, Vorvoreanu, Walker, & Fourney, 2020) and student life (Sillence et al., 2022). This provides a valuable first step for understanding digital hoarding behaviors with respect to personal use digital data.

We draw out four key points for discussion: (i) comparison with workplace settings; (ii) high versus low scorers on the DHQ; (iii) use of smartphone for PIM; and (iv) implications for change and sustainability.

5.1 Comparison with workplace settings

As in previous studies, we note that participants store considerable numbers of digital information items. Sedera et al. (2022) note the average digital data storage volume for an individual is 3.7TB. The average file size has increased more than tenfold since the mid-2000s, and personal collections in general tend to contain fewer documents, but a greater number of large (media) files than work collections (Dinneen & Nyguen, 2021). Our findings, as with previous research, indicate that the range of information item types and the number of items stored on someone's personal device vary considerably. In terms of the number of information items stored, our participants reported having

more items than respondents in studies focused strictly on work-based data (Neave et al, 2019). In a broader, non-work context, our findings sit at the lower end of the items stored range identified by previous research (Dinneen, Julien & Frissen, 2019). This may be because we asked our respondents to focus exclusively on personal use digital data and not to consider work-based data at all. Although it is worth noting that comparisons are not straightforward given the variation in data collection methods and the information item types included.

In terms of information item types, there were more photos and fewer numerical files in personal use contexts compared to work contexts (Neave et al., 2019). The large number of photos stored fits with earlier work on personal versus work based digital hoarding (Sweeten et al., 2018) and resonates with the large volume of photos stored in cloud settings (Brackenbury, McNutt, Chard, Elmore & Ur, 2021). The ease of generating and then often failing to delete multiple copies of rejected photo files adds to the accumulation of this form of digital data (FHE Health, 2020). A recent US survey suggests that 60% of respondents hardly ever deleted photos from their phone (Melore, 2021). This is likely to exacerbate the retrieval difficulties identified in Bergman et al., (2023). Emails continue to be a commonly hoarded information item type as with work settings. However, our respondents stored larger numbers of read and unread emails on their personal devices compared to in workplace settings (Neave et al., 2019). This may relate to more practical and regulatory issues such as workplace rules around email retention, automatic deletion software, and spam filters. It also suggests the underlying motivations or drivers of digital hoarding in personal use settings may vary in comparison with workplace settings with disengagement rather than anxiety or compliance being important (McKellar et al., 2020). Disengagement captures time constraints as well as the problem of not knowing where or how to start deleting and managing files appropriately. Not being able to find emails or other information items leads to frustration and reduced productivity and can leave individuals feeling overwhelmed by their data (Sweeten et al., 2018). We know that anxiety around data can be induced by the visibility of large volumes of data itself. Unread emails, for example are 'visible' and prove stressful for students (Sillence et al., 2022). Taken together these differences suggest the importance of work placed rules around information item deletion and retention (McKellar et al., 2023), differences in methodology and potential differences in how people use local storage versus cloud storage that require further attention.

We have taken a broad approach to digital data and have been influenced by previous studies that have seen participants' own considerations of digital data included (Vitale, Odom, & McGrenere, 2019). Looking at email alongside other information item types, for example, can make comparisons with some PIM studies more complex. However, adopting a user-led approach to studying digital hoarding is important when considering future products and interventions aimed at reducing digital data overload.

5.2 High versus low scorers on the DHQ

High scorers on the DHQ stored more information items than low scorers, and also reported more distress at the thought of losing their items. They were also more attached to their items and found it more difficult to find their information items. Individuals may find it difficult to locate digital content in part due to the fact that collections are often large and deep (Dinneen, Julien & Frissen, 2019). Our findings show that digital hoarders had more difficulty finding their information items although interestingly there was no difference between low and high scorers in terms of the extent to which they organized their items. Although individuals rarely organize information items on their personal devices including their mobile phones (Sedera et al., 2022), it may be that low scorers on the DHQ have less trouble finding what they need and rely on a range of potentially more effective

strategies. In general, little is known about the organization strategies of people with stronger hoarding tendencies in relation to personal digital data. However, the importance of individual differences has been noted in relation to affective responses towards personal use digital data (Alon & Nachmias, 2020) and design concepts around personal information management (Vitale et al., 2019). Understanding the scale of digital hoarding in relation to personal use digital data and the behavioral and affective responses to stored personal use data is an important starting point for the design of suitable interventions to reduce unnecessary digital clutter.

5. 3 Use of smartphone for PIM

The majority of participants chose their smartphone when asked to consider the personal device that contained the majority of their digital information items. Contrary to our expectations, device choice did not impact on the total number of items stored and there were no differences in terms of difficulty finding items on smartphones compared to other devices e.g., laptops. This was unexpected, given previous research indicating that people find data on smartphones easier to discard and often less important (Vitale et al., 2019). Differences in storage capabilities between smartphones and laptops have also been identified previously as contributing to deletion activities (Vitale et al., 2019) although internal storage on smartphones has increased dramatically in the last decade (Steinberg, 2022). Recent work has indicated that information item type might be important when considering device chosen for PIM. We did not find any differences in terms of people reporting difficulties finding items on smartphones and it could be that for certain items, e.g., photos, people are already using technology features to aid retrieval (Bergman et al., 2023). It may also be that specific features of the operating system (OS) on the smartphone contributed in a practical way to information item management behaviors and investigation of the OS in relation to device choice is something to be investigated further. Our sample were relatively young and thus more likely to have high digital competencies and be advanced smartphone users (Wenz & Keusch, 2022) and this will have supported their information item management behaviors on their commonly used device.

The fact that the majority of people chose their smartphone highlights its prominent role in daily life. Indeed, one in five internet users in the UK now access the internet exclusively using a smartphone (Ofcom, 2022). Our findings suggest that in addition to accessing websites and information, users are storing large numbers of information items on their smartphone as well. This raises concerns for data management and security, given that we know compliance with smartphone security behaviors is poor (Shah & Agarwal, 2020) and poorly protected smartphones are likely to be accessed if lost (Egan, 2015). While some people recognize that storing large amounts of unsorted digital data on their personal technology could pose a cybersecurity risk (Sweeten et al., 2018) many individuals are unaware of countermeasures to smartphone loss and theft (Tu, Yuan & Archer, 2014). In terms of PIM, our findings around smartphones raise issues for data management but also point towards the idea of the smartphone as a functional extension of the self, supporting memory for example through stored photos (Park & Kaye, 2019). Interactions between the content and the device in relation to self-extension (Belk, 1988) are worth exploring further. Finally, large volumes of data stored on personal devices suggests the need for increased awareness around digital data hoarding in relation to sustainability issues more broadly (Ovo, 2019).

5.4 Implications for change and sustainability

A study by the Institution of Engineering and Technology (IET), (2021) found that unwanted digital photographs left in storage could accumulate 10.6kg of CO_2 emissions annually for every adult in the

UK. Other 'dirty data' habits included using two or more devices at once, passive streaming and failing to clear archives from messaging services. Hellmich & Dinneen (2022) also stress the importance of deletion for social and environmental sustainability while noting that the scale and structure of file collections stored on devices makes this less straightforward. They argue that the benefits of deletion have been underplayed and reducing data in relation to *inter alia* improved cybersecurity Neave et al., (2020) and negative feelings of overload (Sweeten et al., 2018) are worth promoting.

Examining strategies to encourage responsible personal internet usage, Wathuge and colleagues looked at the different types of motivation that impact upon personal green IT behavior. They noted that different motivations affected utilitarian and hedonic green IT behaviors (Wathuge, Sedera & Sorwar, 2022). While their review did not focus on smartphones specifically, it is worth noting that multiple strategies may be required to change behavior in respect to sustainability in this context.

Individuals vary in relation to their digital data behaviors and in the extent to which they value different solutions designed to assist with data deletion and retention (Vitale et al., 2019). Context and motivation are key in this respect and will have important ramifications for designing effective solutions for addressing digital hoarding. These may include designing appropriate and device specific behavior change interventions around more sustainable digital data use in a personal context.

While further work is needed to explore the role of device in digital hoarding, this study suggests that people accumulate personal use digital data on their devices and have difficulty deleting and managing that data. When designing services and solutions to support data management, it will be critical to be mindful of the extent to which an individual has a digital hoarding tendency. This consideration should extend to various information item types, including emails, which have frequently been studied separately from other information item types, even though they are an integral part of the digital data experience for users.

5.5 Limitations and future work

This is the first study to our knowledge to examine digital hoarding in relation to personal use digital data. Using the validated DHQ in combination with the newly developed Digital Behaviors for Personal use Questionnaire (DBPUQ) has allowed an examination of individuals' digital behaviors regarding personal technology. In terms of study limitations, we hoped that by focusing on a specific device, participants would be more accurate in their reporting of information item numbers and types. While we did ask participants to check on the device before completing the questionnaire, we are unable to verify the accuracy of the figures provided by participants. We followed the self-report data collection approach in Neave et al., (2019) which allowed us to draw comparisons with workplace digital data. The online questionnaire approach was also practical given that the study was conducted during the COVID-19 pandemic. However, the information item management measures we referred to in our questionnaire are all subjective, and it is likely that individuals have different ideas about what counts as well organized. However, subjective feelings around file organization remain important to 'how people feel' about their data (Sillence et al., 2022). Going forward it would be useful to employ a 'guided tour' or 'think aloud' approach (Thomson, 2015) to digital data hoarding, especially in relation to smartphone use, to capture users experience of digital data management.

Exploring age differences in more detail may be useful in relation to understanding anxiety associated with personal use digital data management and PIM strategies more broadly. We know for example, that millennials, those born after 1980, are more anxious about digital hoarding (Sedera et al., 2022) and our sample contained a larger proportion of younger adults and so may not be generalizable to older adults. Future research may also look to investigate age, gender and socio-economic differences in digital hoarding behaviors. There is already an established body of research investigating age and gender differences in cybersecurity behaviors (Branley-Bell et al., 2022), and research suggests that there are also socio-economic differences in smartphone cybersecurity behaviors (Khan et al., 2023).

The range of information item stored on personal devices, in particular smartphones, points to their central role in daily life. Despite many workplace attempts to clearly demarcate work and personal technology through Bring Your Own Device (BYOD) device policies, many people use smartphones for work and personal life accessing work-based email on their personal device and these blurred boundaries would be interesting to explore further in relation to digital hoarding. Likewise, using the DBPUQ to compare behaviors across different devices in more detail including the role of smartphone OS would be a useful next step to understand more about the role of device in digital hoarding. Future qualitative studies could seek to understand how people are using different devices and how the large volume of data stored is impacting on productivity and PIM strategies. As previously detailed, in this study we chose to focus on personal technology believing this still to be the primary way in which people interact with their personal digital data even if it is also stored or backed up to the cloud. Nevertheless, understanding more about how users perceive an integrated approach to digital data management and the role of cloud services would be a useful future line of research. Especially given that we know that cloud deletion preferences are complex and multidimensional (Ramokapane, Such & Rashid, 2022).

6. Conclusion

An examination of individuals' digital behaviors with regard to personal technology has indicated that people store a large number of information item, often on smartphones, suggesting potential problems for PIM strategies and highlighting consequences for mobile phone security. We provide a valuable starting point for understanding digital hoarding in a personal use digital data context, including the types of information items stored, the information item management behaviors and the affective responses to the data. Building on these findings, future work may help to contribute towards the development of suitable design solutions and behavior change interventions aimed at reducing the personal storage of unnecessary digital data.

Notes on contributors

Elizabeth Sillence (elizabeth.sillence@northumbria.ac.uk) is an associate professor in psychology in the Psychology and Communication Technology Lab at Northumbria University. In addition to security and the hoarding of digital data, Liz's research interests centre on eHealth information, examining how people use digital information and advice to inform their health decision-making and improve their wellbeing. ORCID: 0000-0003-1085-7115.

Jordan Dawson (Jordan.a.dawson@northumbria.ac.uk) is a PhD student at Northumbria University and an Associate Lecturer at the Open University. His research areas include computer-human

interactions, internet usage, and internet-based discrimination such as online sexism, homophobia and transphobia. ORCID: 0000-0001-9519-9697

Richard Brown (richard6.brown@northumbria.ac.uk) is a doctoral researcher at Northumbria University. His work investigates a range of health and technology issues using both quantitative and qualitative methods. He has held fellowships with the Parliamentary Office of Science and Technology, the Alan Turing Institute, and currently works as a researcher for the Psychology and Communication Technology Lab at Northumbria. ORCID: 0000-0002-4384-775X

Kerry Lakey (nee McKellar)(kerry.l.lakey@northumbria.ac.uk) is an assistant professor in Psychology at Northumbria University. Her research interests centre around women's health, in particular stigmatised health for example, menstruation, menopause and sexual health. She is particularly interested in how we can use digital technology to benefit health seeking behaviors and behavior change. ORCID: 0000-0002-5112-5300

Nick Neave (nick.neave@northumbria.ac.uk) is a professor in Psychology at Northumbria University. He is Director of the Hoarding Research Group and Chair of the UK Hoarding Partnership, and his research comprise the psychological characteristics of hoarding and collecting behaviors in both physical and virtual realms. ORCID is 0000-0002-5380-4614

References

Ali, I. & Warraich, N.F. (2021).The relationship between mobile self-efficacy and mobile-based personal information management practices: A systematic review. *Library Hi Tech, 39*(1), 126-143.

Ali, I., & Warraich, N. F. (2022). Personal information management through ubiquitous devices: Students' mobile self-efficacy and PIM practices. *Journal of Librarianship and Information Science, 54*(2), 174-187.

Alon, L., Forkosh-Baruch, A., & Nachmias, R. (2020). Gaps in Personal Information Management Behavior of Higher Education Students: Implications for a Training Program. *In EdMedia+ Innovate Learning,* 352-358. Association for the Advancement of Computing in Education (AACE).

Alon, L., & Nachmias, R. (2020). Anxious and frustrated but still competent: Affective aspects of interactions with personal information management. *International Journal of Human-Computer Studies, 144*, 102503.

Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research, 15,* 139-168.

Bergman, O., Gutman, D., & Whittaker, S. (2023). It's too much for us to handle—The effect of smartphone use on long-term retrieval of family photos. *Personal and Ubiquitous Computing, 27*(2), 289-298.

Bergman, O., & Yanai, N. (2018). Personal information retrieval: Smartphones vs. computers, emails vs. files. *Personal and Ubiquitous Computing, 22*(4), 621-632.

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods, 17(*3), 399–417.

Bozacı İ., & Gökdeniza, İ. (2020). Development of a digital photo hoarding scale: A research with undergraduate students. *Management Science Letters, 10,* 2193-2200.

Brackenbury, W., McNutt, A., Chard, K., Elmore, A., & Ur, B. (2021, October). KondoCloud: Improving information management in cloud storage via recommendations based on file similarity. *In The 34th Annual ACM Symposium on User Interface Software and Technology*, 69-83.

Branley-Bell, D., Coventry, L., Dixon, M., Joinson, A., & Briggs, P. (2022). Exploring age and gender differences in ICT cybersecurity behavior. *Human Behavior and Emerging Technologies, 2022,* 2693080.

Creegan, T.D. (2017). A Study of the Personal Information Management Practices of Librarians. Victoria University of Wellington. Available at: http://researcharchive.vuw. ac.nz/xmlui/bitstream/handle/10063/6625/paper_access. pdf?sequence=1

Cushing, A. L. (2011). Self extension and the desire to preserve digital possessions. *Proceedings of the American Society for Information Science and Technology, 48*(1), 1-3.

Dinneen, J. D., & Julien, C. A. (2019). What's in people's digital file collections? *Proceedings of the Association for Information Science and Technology, 56*(1), 68-77.

Dinneen, J. D., Julien, C. A., & Frissen, I. (2019, May). The scale and structure of personal file collections. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Dinneen, J. D., & Krtalić, M. (2020). E-mail as legacy: managing and preserving e-mail as a collection. portal: *Libraries and the Academy, 20*(3), 413-424.

Dinneen, J. D., & Nguyen, B. X. (2021). How Big Are Peoples' Computer Files? File Size Distributions Among User-managed Collections. *Proceedings of the Association for Information Science and Technology, 58(1*), 425-429.

DiStefano, Christine; Zhu, Min; and Mîndrilã, Diana (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research, and Evaluation, 14* , Article 20.

Egan, G. (2015). The Honey Stick Project Revisited: How Secure Are Your Mobile Devices? Available at: https://www.proofpoint.com/us/security-awareness/post/honey-stick-project-revisited-how-secure-are-your-mobile-devices

FHE Health (2020). Body Image and Social Media Questionnaire https://fherehab.com/news/bodypositive/

Gormley, C. J. & S. J. Gormley. (2012). Data hoarding and Information Clutter: The Impact on Cost, Life Span of Data, Effectiveness, Sharing, Productivity, and Knowledge Management Culture. *Issues in Information Systems, 13*(2), 90-95.

Gulotta, R., Odom, W., Forlizzi, J., & Faste, H. (2013, April). Digital artifacts as legacy: exploring the lifespan and value of digital data. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1813-1822).

Hellmich, H., & Dinneen, J. D. (2022). Making space for the future: The importance of deletion for LIS and the information society. *ISIC `22: Proceedings of the Information Seeking in Context (ISIC) Conference.*

Henderson, S. (2011). Document duplication: How users (struggle to) manage file copies and versions. *Proceedings of the American Society for Information Science and Technology, 48(1),* 1– 10.

IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.

Kaye, J. J., Vertesi, J., Avery, S., Dafoe, A., David, S., Onaga, L., Rosero I. & Pinch, T. (2006, April). To have and to hold: exploring the personal archive. *In Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 275-284).

Khan, N. F., Ikram, N., & Saleem, S. (2023). Digital divide and socio-economic differences in smartphone information security behavior among university students: empirical evidence from Pakistan. *International Journal of Mobile Communications, 22(1),* 1-24.

Krtalić, M., Dinneen, J. D., Liew, C. L., & Goulding, A. (2021). Personal collections and personal information management in the family context. *Library Trends, 70*(2), 149-179.

Neave, N., Briggs, P., McKellar, K., & Sillence. E. (2019). Digital Hoarding Behaviors: Measurement and Evaluation. *Computers in Human Behavior, 96*, 72–77.

Neave, N., McKellar, K., Sillence, E., & Briggs, P. (2020). Digital hoarding behaviors: Implications for cybersecurity. *In Cyber influence and cognitive threats* (pp. 77-95). Academic Press.

McKellar, K., Sillence, E., Neave, N., & Briggs, P. (2020). There Is More Than One Type of Hoarder: Collecting, Managing and Hoarding Digital Data in the Workplace. *Interacting with Computers 32*(3), 209–220.

McKellar, K., Sillence, E., Neave, N., & Briggs, P. (2023). Digital accumulation behaviors and information management in the workplace: exploring the tensions between digital data hoarding, organisational culture and policy. *Behavior & Information Technology*, 1-13.

Melore, C. (2021). Digital hoarders: 3 in 5 people hardly ever delete photos from their smartphones.

Available at: https://studyfinds.org/digital-hoarders-rarely-delete-photos-from-smartphone/

Ofcom (2022). Adults' Media Use and Attitudes report. https://www.ofcom.org.uk/__data/assets/pdf_file/0020/234362/adults-media-use-and-attitudes-report-2022.pdf

Oravec, J. A. 2018. Digital (or Virtual) Hoarding: Emerging Implications of Digital Hoarding for Computing, Psychology, and Organization Science. *International Journal of Computers in Clinical Practice 3*(1), 27–39.

Ovo (2019). 'Think Before You Thank': If every Brit sent one less thank you email a day, we would save 16,433 tonnes of carbon a year - the same as 81,152 flights to Madrid. Available at: https://www.ovoenergy.com/ovo-newsroom/press-releases/2019/november/think-before-you-thank-if-every-brit-sent-one-less-thank-you-email-a-day-we-would-save-16433-tonnes-of-carbon-a-year-the-same-as-81152-flights-to-madrid

Park, C. S., & Kaye, B. K. (2019). Smartphone and self-extension: Functionally, anthropomorphically, and ontologically extending self via the smartphone. *Mobile Media & Communication, 7*(2), 215-231.

Peterson, R. A., & Peterson, M. R. A. (2020). Package 'bestNormalize'. *Normalizing Transformation Functions. R package version*, *1*(1).

Ramokapane, K. M., Such, J., & Rashid, A. (2022). What Users Want From Cloud Deletion and the Information They Need: A Participatory Action Study. *ACM Transactions on Privacy and Security, 26(*1), 1-34.

R Core. (2021). Core R: a language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna*.

Sannon, S., Vorvoreanu, M., Walker, K., & Fourney, A. (2020). " Am I doing this all wrong?" Challenges and Opportunities in Family Information Management. *Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2),* 1-28.

Sedera, D., & Lokuge, S. (2018). Is digital hoarding a mental disorder? Development of a construct for digital hoarding for future IS research. *In Proceedings of the 39th International Conference on Information Systems (ICIS 2018).* Association for Information Systems (AIS).

Sedera, D., Lokuge, S., & Grover, V. (2022). Modern-day hoarding: A model for understanding and measuring digital hoarding. *Information & Management, 59*(8), 103700.

Shah, P., & Agarwal, A. (2020). Cybersecurity behavior of smartphone users in India: an empirical analysis. *Information & Computer Security, 28*(2), 293-318.

Sillence, E., Dawson, J. A., McKellar, K., & Neave, N. (2022). How do students use digital technology to manage their university-based data: strategies, accumulation difficulties and feelings of overload? *Behavior & Information Technology, 42*(14), 2442-2451

Steinberg, N. (2022). How Much Storage (in GB) Do I Need In My Phone? https://www.lifewire.com/how-much-phone-storage-5272110 (accessed 17th July, 2023).

Subedi, B. P. (2016). Using Likert type data in social science research: Confusion, issues and challenges. *International Journal of Contemporary Applied Sciences, 3*(2), 36-49.

Sweeten, G., Sillence, E., & Neave, N. (2018). Digital hoarding behaviors: Underlying motivations and potential negative consequences. *Computers in Human Behavior, 85,* 54-60.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273-1296.

The Institution of Engineering and Technology (2021). Dirty Data. https://www.theiet.org/media/press-releases/press-releases-2021/press-releases-2021-october-december/26-october-2021-dirty-data/ (accessed 17th July, 2023).

Tu, Z., Yuan, Y., & Archer, N. (2014). Understanding user behavior in coping with security threats of mobile device loss and theft. *International Journal of Mobile Communications, 12*(6), 603-623.

Tugtekin, E. B. (2022). Investigation of the relationship between digital hoarding, information technologies self efficacy and anxiety. *European Journal of Education Studies, 9*(1).157-177.

Thomas, L., & Briggs, P. (2016). Reminiscence through the lens of social media. *Frontiers in Psychology, 7*, 870.

Thomson, L. (2015). The guided tour technique in information science: explained and illustrated. *In ASIS&T '15: Proceedings of the 78th ASIS&T Annual Meeting of the Associations for Information Science and Technology*. St. Louis, Missouri, 135:1–135:5

Vitale, F., Janzen, I., & McGrenere, J. (2018, April). Hoarding and minimalism: Tendencies in digital data preservation. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Vitale, F., Odom, W., & McGrenere, J. (2019, June). Keeping and discarding personal data: exploring a design space. *In Proceedings of the 2019 on designing interactive systems conference* (pp. 1463-1477).

Wathuge, A., Sedera, D., & Sorwar, G. (2022). Greening the Personal Internet Use: Findings from a literature review. *In Australasian Conference on Information Systems* (pp. 1-13). Association of Information Systems.

Wenz, A., & Keusch, F. (2022). The second-level smartphone divide: A typology of smartphone use based on frequency of use, skills, and types of activities. *Mobile Media & Communication, 11*(3), 459-483.

Zhang, P. & Liu, C. (2015). Personal information management practices of Chinese college students on their smartphones. *In Proceedings of the third international symposium of Chinese (CH*I), Seoul, South Korea, 18–19 April, pp. 47– 51. New York: ACM.

Zimmerman, D. W., & Zumbo, B. D. (1993). Rank transformations and the power of the Student t test and Welch t'test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 47*(3), 523. Chicago

Appendix

The Digital Hoarding Questionnaire (DHQ) and the Digital Behaviors for Personal use Questionnaire (DBPUQ)

We now gather and use many different types of digital information with different formats and on different devices. We are interested in how you behave towards, and how you think about the different types of digital information you typically use in your personal (**not** working life).

Please answer the following questions as honestly as you can, if you prefer not to give an answer then leave it blank.

PART 1: ABOUT YOU

1.1 I am: male ☐ female ☐ prefer not to say ☐

1.2 I am ____ years old

PART 2: ABOUT YOUR PERSONAL DIGITAL DEVICES

In your personal life, how many devices where digital information might be stored or accessed (e.g. phone, camera, laptop etc) do you possess? In the list below state how many of them you currently own:

TYPE OF DIGITAL DEVICE Add any that we may have missed out	**HOW MANY OF THESE DO YOU HAVE?** Insert a number, if you have none put '0'
Mobile phone	
Laptop	
Personal Computer (PC)	
eReader	
Camera	
Tablet computer e.g. iPAD	
External hard drive	
USB drive	

PART 3: HOW YOU FEEL ABOUT DIGITAL INFORMATION IN GENERAL

We are interested in how people feel about digital materials in their personal life. These materials which we refer to as 'files', include emails, email attachments, photos, music files, movie files, apps, PDF's, etc. When you answer, do not consider spam/junk files, which many people delete instantly.

Please answer the following statements by selecting the most appropriate number, where 1 = not at all to 7 = very much so

	Not at all						*Very much so*
3.1. I find it extremely difficult to delete old or unused files	1	2	3	4	5	6	7
3.2. I tend to accumulate digital files, even when they are not directly relevant to me	1	2	3	4	5	6	7
3.3. Deleting certain files would be like deleting a loved one	1	2	3	4	5	6	7
3.4. If I delete certain files I feel apprehensive about it afterwards	1	2	3	4	5	6	7
3.5. I strongly resist having to delete certain files	1	2	3	4	5	6	7
3.6. I feel strongly that some files might be useful one day	1	2	3	4	5	6	7
3.7. I lose track of how many digital files I possess	1	2	3	4	5	6	7
3.8. Deleting certain files would be like losing part of myself	1	2	3	4	5	6	7
3.9. Thinking about deleting certain files causes me some emotional discomfort	1	2	3	4	5	6	7
3.10. At times I find it difficult to find certain files because I have so many	1	2	3	4	5	6	7

DIGITAL BEHAVIORS FOR PERSONAL USE QUESTIONNAIRE (DBPUQ)

PART 4: ABOUT THE NUMBER OF YOUR PERSONAL DIGITAL FILES

We are now interested in the range of digital materials that people typically have access to in their personal life, how many of these materials people typically possess, and how they behave towards them. We will refer to these materials as 'files', these include photos, emails, email attachments, apps, music files etc. When you answer, do not consider spam/junk files and only think about your most commonly used email account.

Choose a personal device that contains the majority of your digital files – this might be your phone, your laptop, your computer etc. This might not be the device that you use the most often (e.g. a phone), but the device that contains the majority of your personal digital files (not files that are for work, study etc.). Please do not consider cloud storage.

What device have you chosen? _____

Below is a list of common digital items you might currently have stored on this device. For each one, please indicate approximately how many you have right now. If you have access to this device please provide an exact number, if you do not have access please try to give an accurate estimate of the number of files you have.

TYPE OF DIGITAL FILE (please add any we might have missed)	EXACT OR APPROXIMATE NUMBER (if none put 0)
Photographs	
Music files	
Apps	
Videos	
Movies	
e-books	
Read emails currently in in-box	
Unread emails currently in in-box	
Emails currently in 'deleted' folder	
Emails in saved/archived folders:	
Text files *(e.g. word documents, PDFs etc)*	
Numerical files *(e.g. spreadsheets, databases etc)*	

PART 5: ABOUT YOUR DELETION BEHAVIORS

Typically, how often do you tend to delete the following types of digital files? When you answer, do not consider spam/junk files.

Please tick one box that best describes your deletion habits for each file type.

File Type (Add any in we might have missed)	**I typically delete these daily**	**I typically delete these weekly**	**I typically delete these monthly**	**I typically delete these yearly**	**I hardly ever delete these files**
Photographs					
Music files					
Apps					
Videos					
Movies					
e-books					
Read emails currently in in-box					
Unread emails currently in in-box					
Emails currently in 'deleted' folder					
Emails in saved folders					
Text files *(e.g. word documents, PDFs etc)*.					
Numerical files *(e.g. spreadsheets, databases etc)*.					

PART 6: HOW YOU FEEL ABOUT YOUR PERSONAL DIGITAL FILES

In the sections above you have told us about the type and number of digital files you possess. Now, for each file type think about how attached you are or not to this specific type of file. Then think how you would feel if you lost these files and could not recover them.

For each file type tell us how strongly attached to them you are, and how strongly you would feel if they were deleted/lost forever.

Use the scale where 1 = not at all and 7 = very much so

File Type (Add any in we might have missed)	**To what extent do you feel attached to this file type?**	**To what extent would you feel distress if this file type were lost for good?**
Photographs		
Music files		
Apps		
Videos		
Movies		
e-books		
Emails		
Text files *(e.g. word documents, PDFs etc)*.		
Numerical files *(e.g. spreadsheets, databases etc)*.		

PART 7 HOW I ORGANISE MY PERSONAL DIGITAL FILES

For each of the file types we have considered so far, please tell us how you typically organise those files, and the difficulty you might have in finding them easily again.

For the first question use the scale where 1 = no organisation at all to 7 = very well organized. For the second question use the scale where 1 = no difficulty at all to 7 = very difficult indeed.

File Type (Add any in we might have missed)	**To what extent do you organise these files?**	**To what extent do you have difficulty in finding these files again?**
Photographs		
Music files		
Apps		
Videos		
Movies		
e-books		
Emails		
Text files *(e.g. word documents, PDFs etc)*.		
Numerical files *(e.g. spreadsheets, databases etc)*.		