There is more than one type of hoarder: collecting, managing, and hoarding digital data in the workplace

ABSTRACT

There are organizational and individual problems associated with the excessive accumulation of digital material, but little is known about why people hoard digital information in the workplace. We interviewed 20 participants from two large knowledge-intensive organizations (one academic, one commercial). These participants scored highly on the Digital Hoarding Questionnaire (DHQ). We asked them to discuss their information management practices, their reasons for keeping large amounts of digital data and the difficulties they faced in relation to deleting data. Using thematic analysis, we identified four underlying dimensions of digital hoarding (anxiety, disengagement, compliance and collection). We discuss these in relation to the consequences of digital hoarding for individuals and organizations - considering the implications for organizational culture, digital ownership and cybersecurity.

KEY WORDS:

Data management, digital hoarding; workplace

Research highlights:

- Digital hoarding in the workplace is common but little is currently known about why people hoard digital information.
- Using the published Digital Hoarding Questionnaire we identified people who displayed hoarding tendencies and interviewed them about their information management, retention and deleting difficulties.
- Four underlying dimensions of digital hoarding were identified; anxiety, disengagement, compliance and collection.
- This is the first study to identify underlying dimensions of digital hoarding and their threat to the workplace.

INTRODUCTION

The hoarding of digital data has only recently been identified as an important research topic (Oravec, 2017, 2018; Vitale, Janzen & McGrenere, 2018). Recent research has gone on to explore some of the technical and organisational implications of data hoarding. Vitale et al., (2018), for example, suggested ways that technology might better support both people who display hoarding (keeping most data) and minimalists (keeping as little data as possible) tendencies, while Sweeten, Sillence & Neave (2018) noted the potential costs of failure to delete unnecessary digital files. These costs included reduced productivity, stress and anxiety and cybersecurity threats. Both these papers addressed digital hoarding tendencies in the general population and examined personal information management strategies across a range of settings and devices.

For organizations, excessive data hoarding can be particularly problematic. Data clutter is highly prevalent and the impact of data hoarding on a business can be large, with cost, data lifespan, effectiveness, productivity and knowledge management all affected (Gormley & Gormley 2012). Digital data hoarding also has a cybersecurity cost, as the stored material can be used by malicious insiders or by external agencies involved in data theft or industrial espionage. The storage and management of digital data is thus a topic of interest for researchers and practitioners alike.

From a design perspective, researchers have sought to understand how people acquire, store and curate personal data for the purposes of reminiscence or legacy creation (Odom, Zimmerman & Forlizzi, 2011; van Gennip, 2015; Thomas & Briggs 2014; Gulotta, Odom, Forlizzi & Faste, 2013; Gerritsen, Tasse, Olsen et al., 2016; Watkins, Sellen & Lindley, 2015). People generally see their digital possessions as an extension of the self and become increasingly attached to them (Cushing, 2011, 2013).

From an organisational perspective, data curation cycles of acquiring, storing, processing and reusing data may be affected by both individual and organizational factors (Whittaker, 2011). For example, people use different strategies to manage emails and other digital documents and to varying extents adopt a filing, piling or structuring approach to digital data management (Henderson, & Srinivasan, 2011). *Filers*, for example, are relatively organised and save documents, emails and bookmarks into folders while *Pilers* or *No Filers* are more likely to scan and retrieve emails or search for documents rather than save them to a folder (Whittaker & Sidner, 1996). *Structuring Filers* (Henderson, & Srinivasan, 2011) create folders that enable easy retrieval of documents but also provide a conceptual overview of the data stored. Such personal information management (PIM) strategies can be linked to personality characteristics of the individual particularly the traits of conscientiousness and neuroticism (Massey, TenBrook, Tatum & Whittaker, 2014), In the Massey et al., study, more

conscientious individuals were better organized and more neurotic individuals kept greater amounts of information to hand on their desktop, in response to work pressure.

It is clear that many people have a strong relationship with their data. For example, a study designed to understand the archiving strategies of academics (both physical and digital) found that people kept and curated material to build a legacy, share information, preserve important possessions, and construct an identity (Kaye, Vertesi, David, Onaga & Pinch, 2006). More recently, Vitale et al. (2018) showed that the digital data behaviours of both people who display hoarding and minimalist tendencies include both practical and emotional elements. Yet to date, the overwhelming majority of papers on PIM have focused on the practicalities of managing data at scale.

There is a very small but nascent body of research on digital or virtual hoarding (Oravec, 2015, 2017, Luxon, Hamilton, Bates & Chasson, 2019). In part, this has been triggered by a stronger recognition of the psychological and lifestyle problems associated with *physical* hoarding, but it has encouraged researchers to examine the ways in which people accumulate data at scale and/or struggle to delete digital data files. Defined as *"…the acquisition of and failure to discard digital content, leading to the accumulation of digital clutter"* (Sedera & Lokuge, 2018), a small number of recent studies have identified the negative consequences of digital hoarding, including feelings of anxiety in relation to the accumulation of data (van Bennekom et al., 2015; Sedera & Lokuge, 2018; Sweeten et al., 2018; Vitale et al., 2018).

Only a few studies have asked people to describe their digital hoarding behaviors in any detail. Vitale et al., (2018) conducted a qualitative investigation of both hoarding and minimalist data storing practices. Participants were asked to demonstrate their data storage practices using their own devices and to discuss the value of their digital data. Sweeten et al. (2018) carried out a qualitative study of digital hoarding behaviors, motivations and consequences in 45 individuals and identified themes common to physical hoarding. The study identified five main barriers to data deletion: 1 - keeping it for the future/or just in case; 2 - keeping it as evidence; 3 - too time consuming or just too lazy to delete it; 4 - emotional attachment; and 5 - not being my problem to delete it. Both of these studies asked the general population about their data management practices across a range of settings and devices, and neither focused specifically on digital hoarding behaviors within workplace settings.

Yet workplace settings provide an interesting context to study digital hoarding. As digital hoarding rises, businesses find it more difficult to extract value from the stored information and the risks associated with that information grow significantly (CGOC 2017). These risks have increased recently, with new privacy and data protection legislation that regulates the storage of personal data such as the General Data Protection Regulation (GDPR) in Europe which came into force on May 25th 2018

(Information Commissioners Office, 2018). GDPR is designed to harmonize data protection law across Europe, and to bring the law up to date with technological advancements around the increasing use of digital data. In the wake of GDPR legislation, both individuals and organizations could unwittingly store data illegally. Similar laws in the US e.g. the California Consumer Privacy Act (CCPA) (State of California Department of Justice, 2020) mean this is a timely issue, with relevance beyond Europe. In addition, there are cybersecurity risks associated with hoarding behaviors, given that the stored material could be mined for social engineering attacks, or used by disgruntled employees who have at their disposal a repository of confidential or possibly embarrassing material that may date back many years.

A number of individual and organisational factors influence the extent to which people hoard data (Oravec, 2017). For example, employees may be prone to digital hoarding when swept up in an organisational change programme. There may be increased knowledge hoarding in times of uncertainty, or when there is a sense of lost trust within the organization (Holten et al., 2016). During such times, individuals accumulate organizational information in strategic ways, holding onto information that may be of future use to themselves and thereby creating a false sense of 'uncertainty avoidance' (Evans, Hendron & Oldroyd, 2014; Gormley & Gormley, 2012).

Study aims*

In our study, we focused on knowledge-intensive workplace data (e.g. research organizations) and drew specifically on a population of people who scored highly on the DHQ (as defined by their scores on the Digital Hoarding Questionnaire: DHQ; Neave, Briggs, McKellar & Sillence, 2019). The aim was to understand how these individuals manage their workplace data; specifically their motivations and strategies for accumulating and keeping large amounts of work based digital data. We also sought to understand the barriers to change with respect to digital hoarding. It is important to note here that we did not make *a priori* assumptions regarding the extent to which our participants were aware of or engaged with the term 'digital hoarding' and we remained mindful of the societal stigma associated with physical hoarding. Furthermore, we drew on related literatures that distinguished between hoarders and collectors to draw a more nuanced account of these accumulation and deletion behaviors. Kibby (2009) for example, described how the acts of accessing, tagging, and storing digital music files can instil meaning into these virtual objects and turn the files into a "collection" rather than just a hoard (p. 428). Belk, Wallendorf & Sherry (1988) also posits that that 'collectors' are distinct from hoarders in that their collections have specific meanings and structures.

METHOD

Participants

The digital hoarding questionnaire DHQ (Neave et al., 2019), an 11-item scale that measures digital hoarding behaviors in the workplace, was sent to two large organizations in the North East of England, an academic organization (>1000 employees) and a commercial organization (>800 employees). These two organizations were chosen because their employees typically deal with large amounts of digital data. Academics have also previously provided a useful focus for studies on data and information management (Kaye et al., 2006). The organizations differed in terms of their local policies regarding data storage. The commercial organization had specific policies regarding cloud services and shared access to files while the academic organization allowed employees to use cloud storage but was not prescriptive in this respect.

The DHQ provides scores on two factors – difficulty discarding and accumulation. A total of 111 participants completed the questionnaire (79 from the academic organization and 32 from the commercial organization), and 50 participants left their email address at the end of the survey indicating that they were willing to take part in a follow up focus group. Of the 50 participants remaining (35 from the academic organization and 15 from the commercial organization), 47 scored highly on the DHQ (i.e. had scores over 16 for difficulty discarding and 15 for accumulating, (for more information about the scoring see Neave et al. 2019), and all were contacted for the follow up study. There were no significant differences between the two organizations for difficulty deleting or accumulating scores. Of those 47 contacted, we recruited 20 of these high scorers to take part in the focus groups. The twenty final participants (15 females, 5 males) ranged from 18 to 59 years of age (mean: 34.6 years SD: 10.05 years). Everyone who took part was either in full-time or part-time employment, used a computer as part of their job and had been with their company for at least 6 months. Table 1 (below) provides an overview of each participant's professional job title and digital hoarding scores (taken from the DHQ). Participants were each compensated a £10 Amazon gift voucher for taking part.

Table 1: Overview of participants' demographic information and digital hoarding scores.

Participant – job title	Organization	DBQ Score (Difficulty discarding/ accumulating)	Data protection responsibility
P1: Senior Researcher	Academic organization	DD:25 A:20	Yes
P2: Researcher	Academic organization	DD:20 A:18	Yes
P3: Demonstrator (PhD)	Academic organization	DD:17 A:16	Yes
P4: Demonstrator (PhD)	Academic organization	DD:25 A:24	Yes
P5: Faculty Associate	Academic organization	DD:20 A:18	Yes
P6: Senior Lecturer	Academic organization	DD:16 A: 20	No

P7: Senior Lecturer	Academic organization	DD:17 A:20	Yes
P8: Senior Lecturer	Academic organization	DD:23 A:22	Yes
P9: Senior Researcher	Academic organization	DD:20 A:16	Yes
P10: Demonstrator (PhD)	Academic organization	DD:28 A:19	Yes
P11: Demonstrator (PhD)	Academic organization	DD:16 A:20	Yes
P12: Associate Professor	Academic organization	DD:22 A:36	Yes
P13: Senior Lecturer	Academic organization	DD:28 A:32	Yes
P14: Demonstrator	Academic organization	DD:24 A:28	Yes
P15: Senior Lecturer	Academic organization	DD:19 A: 16	Yes
P16: Demonstrator (PhD)	Academic organization	DD:20 A:18	Yes
P17: Research Chemist	Commercial organization	DD:20 A: 18	No
P18: Senior Scientist	Commercial organization	DD:18 A:24	No
P19: Intern Scientist	Commercial organization	DD:20 A:18	No
P20: Research Scientist	Commercial organization	DD:22 A:23	Yes

Materials and procedure

Following ethical approval from (blank for review), the DHQ was sent to all staff in the two organizations. Individuals who scored highly were contacted and invited to take part in a focus group. A focus group approach was taken after initial pilot work indicated that work practices around digital data management were best discussed within group settings. Participants helped each other develop a common and more nuanced understanding of 'digital data' and were more willing to share in an open and more informal setting.

The focus group schedule took as it starting point the questions devised by Sweeten et al., (2018). The questions were developed and extended to include different types of digital data while keeping a focus on workplace settings. The questions were devised to explore how individuals managed their digital files focusing on their work computer and to examine their motivations for digital hoarding. Importantly, the term 'digital hoarding' was only mentioned towards the end of the discussion to gauge perceptions and attitudes towards the concept.

Example questions on the interview schedule included: *"Tell me what you typically do when an email lands in your inbox?", "Thinking about the emails in your inbox, what proportion of those emails are valuable to you right now or in the future?"* and *"To what extent do you have difficulties discarding files that are no longer relevant?"*

One member of the research team conducted all of the focus groups, and there were 2-5 participants in each focus group. Each focus group took place in a quiet location at the participants' workplace. The average length of the focus groups was 45 minutes. All interviews were digitally recorded and later transcribed verbatim. Participants were informed about the confidentiality procedures in place, how their data was to be used and that they were free to withdraw from the study at any time without

explanation. All participants were provided with an information sheet, signed an informed consent form and were fully debriefed at the end of the session.

Analysis procedure

The first two authors [KM, ES] analysed the data using Braun & Clarke's (2006) approach to thematic analysis. First, the data from all of the focus groups was transcribed verbatim and initial ideas and thoughts were noted down, the transcripts were then read and re-read several times. We began by examining elements of the transcripts that contained evidence of motivational and explanatory accounts of digital hoarding practices by our participants. We took an inductive approach and generated sets of initial codes that represented the different ways in which people accounted for or simply described their hoarding. We found that some initial codes could be placed under more than one category. We kept the codes as close to the original data as possible and grouped initial codes semantically. These groupings were incorporated into subthemes and these were categorized into four underlying dimensions that made coherent sense and captured the different motivations for hoarding behavior that we saw in the data. We then reviewed the data with the rest of the research team [NN, PB] to ensure that these motivational categories were a good fit with the data including checking for negative cases. During this phase of analysis, any disagreements were resolved through discussion and we finally worked collectively to define and name each of the four dimensions and sought out example quotes from the transcripts to illustrate each dimension type.

We then took a separate, deductive approach and examined the data to identify instances when participants indicated reluctance or difficulty in terms of making changes to their current practices. We coded these instances according to the concepts identified in the small body of literature on barriers to data deletion (Vitale et al., 2018, Sweeten et al., 2018) but we were also open to identifying new codes within the data This analysis was performed by [KM and ES] and then checked with the remaining research team [NN, PB]. Overall, the aim of our analysis was to provide a clear account of the ways in which digital hoarding behaviors are driven, justified and understood by our participants.

RESULTS

We identified four underlying dimensions of digital hoarding in the workplace. *'Anxiety, 'Compliance', 'Disengagement' and 'Collection'*. The findings are presented below followed by a discussion of the results.

Anxiety driven digital hoarding

Anxiety drove many of our participants to keep excessive amounts of digital data. They were nervous about deleting any digital data and believed that they might need their files again in the future. They perceived some value to the digital files and if given the opportunity would prefer to be able to organize their files rather than have to delete them.

> *I have got a feeling like, you know, there is something, you know maybe I am working on something and I think, in the future I might need to, I might need to refer back to this document that might be relevant, then I would be more likely to kind of cling onto it (P16, Demonstrator/PhD)*

Participants recognized that many of the emails they kept were no longer valuable but were still reluctant to delete them in case they were needed in the future or provided a source of evidence.

> *A lot of emails that I have kept, so, even stuff like confirming maternity leave, when I was on maternity leave, so it's just like, anything to do with HR, I am always very keen to keep a record of and make sure It is in a folder somewhere. I mean, I will never need that email saying you are on maternity leave from month A to month B, it is just in case [..] I suppose it is the security of knowing that you have got it. (P1, Senior Researcher)*

Keeping data even if its value is unclear to participants provides a level of security and comfort for participants. Participants were also anxious to preserve information that might become important at some future point in time – deciding that the best plan was to keep everything, just in case. This meant that deletions were limited to item such as spam emails i.e. information that was clearly neither personally associated with them nor directly relevant to their job.

> *Yeah, the only things I ever delete, is when I am on a mailing list and it is just kind of a, you know, news and updates from a company or something, that is the only thing I would delete, I never, I never feel like I want to delete anything else, just in case something comes up and I have to remind myself about something (P10, Demonstrator/PhD)*

Anxiety provoked a sense of ownership over emails and digital files for some participants such that, even though they were happy to share some documents, they viewed many of the files as their own. Anxiety also led some participants to keep digital files on their work computer but also have files on the cloud or external drives. Participants' perceived sense of ownership meant they would want to take their files and emails with them if they were to leave their job.

> *The panic that you think you might need them, but I don't, I probably would just take my PhD stuff, like my research, or the nursery files we could take those (P7, Senior Lecturer)*

Compliance driven digital hoarding

For some participants, storing digital files and emails on work computers was driven by compliance. These participants described having a good knowledge of the emails and files they currently have stored and kept the data in order to comply with their company's policies. The digital data was often well organised and they were likely to keep digital data that had been sent to them by their managers or colleagues.

Where the hoarding behavior is underpinned by compliance, participants perceive little personal value in the digital data they have stored and believe the digital data is for their company's benefit rather than their own. One of the reasons they keep all this data is in case their company was audited. To be compliant, these participants were happy with the amount of digital data they had stored, and believed that it was evidence of them working hard, doing the right thing and following the rules.

> *I have literally got my work career is 20 years of, deleted emails that aren't there, and I, now wouldn't really, I am glad I don't have them, but they are useful for the companies that I worked for and you know that, that is for them if they ever need to audit, any legal decisions or anything that I have made at certain points in time, that is where they will be useful*, *(P13, Senior Lecturer)*

Once the data is no longer required, these participants were also happy to delete information. Participants believed they needed to keep all this digital data in order to be complaint and so did not feel anxious or worried about the amount of information that they had. Likewise, they felt no anxiety over discarding information, as long as they had been 'given permission' to delete it. Deletion behaviours may occur, for example following the completion of a specific project or set of tasks as in the example below.

> *I would keep it all in one file in my inbox, and, once that is all done and dusted, then I get rid of it.* (P15, Senior Lecturer)

In other cases, participants were happy to delete data once their colleagues or managers had made it clear that it was no longer needed. In both cases, it was clear that deleting the data once it was no longer required was easy to do for these participants as the 'quite happy' indicates in the example below

> *If I don't need it for anything, then yeah, quite happy to press the delete button and I do try to do that (P20, Research Scientist)*

Where hoarding behavior is based on compliance, participants had a reduced sense of ownership of the data they had stored and felt happy to share their information with other people. Data was kept

on their work computer rather than external or cloud storage, unless they had a shared cloud system at their company. These participants kept a lot of their information on shared drives, as they were happy to share data with other people. They did not see the data as their own personal data but rather data they kept for the company.

> *See, again I will be different from you guys because it is collaborative work, responses, so it is their files really, and it is my PI files, and it Is my colleagues working with me, it is our files, (P9, Senior Research Assistant)*

Different organizations, of course, promote and encourage different working practices for their employees and provide individuals with more or less discretion around data sharing and data storage. For some of the participants in our study, keeping data was seen as part of doing their job and being efficient in how they managed their work. While hoarding is typically associated with individual factors, hoarding that is driven by compliance may well still lead to sharing behaviors especially in situations involving team or collaborative working environments.

Disengagement driven digital hoarding

Disengagement characterized many of our participants hoarding behaviors. Our participants had a lot of digital data stored in their email inboxes and in their folders. However, this data was not well organized. Being disengaged with data management behaviors meant that participants felt they lacked control over the digital data they had acquired. Files were often sent to them unrequested, and over time, these files had slowly accumulated almost without the participant realizing.

> *I am just looking at mine and I am just realising what a complete mess it is in. I have got all of these folders with particular, you know, projects and all this kind of stuff, and then I have got one that I have just called read, because I can't decide what folder to put it in, and I have looked in the read and there is 3052 just in the read and what are they doing. You know, so, it is a complete shambles. (*P5, Faculty Associate)

Where data accumulation had been driven by a general sense of disengagement, many participants felt that it was now too late to try to organize the data and they should have started good habits around data management much earlier. In addition, some participants admitted they were just too lazy or did not have the time to delete the information and again this reflects the passive disengagement that drives their hoarding behaviour. For these participants, the data was not seen as being particularly valuable but nor is the hoarding behavior seen as problematic. Participants were not concerned about the amount of digital data they had only that it would simply take too long to delete the data.

It is because I am lazy, if I will be honest, but I just think the effort that I have to go to delete it, as oppose to close that notification and continue doing what I am doing, I don't see much point in it. (P14, Demonstrator)

These participants had a reduced sense of ownership over the data they had stored and felt that much of it had been accumulated over time and almost accidentally. Deleting this data would not be problematic as little of it felt 'like theirs' although deleting data was unlikely to occur because of the time costs involved.

Some of them, I mean there is an awful lot of stuff that is just, yeah, I think, yeah, I think there is a mixture of stuff, where I am like yeah because those were to me specifically so those are like my emails, but then there is more kind of generic stuff, where you are just kind of like, but is potentially useful information, and I might kind of need that in the future, but it is not kind of my email (P4, PhD/Demonstrator)

The concept that the 'moment has passed' for organizing digital data was a common idea in the interviews. Likewise, simply being too lazy to start looking through the data or feeling unconcerned about the volume of data were also ideas that participants discussed.

Collection driven digital hoarding

For some participants, hoarding large amounts of digital data was something that was a purposeful decision. Where collection drove the digital hoarding, participants had made the decision to store and not to delete data. For example, participants had large amounts of emails stored in their inbox and used them as prompts to action. These emails were typically systematically organized into folders, and were rarely deleted. The only emails considered for deletion were obvious spam emails (for example, phishing emails). Again, participants felt that many files, regardless of 'how old they were' had the potential to be valuable and perhaps needed again in the future

Yeah those old files are surprisingly still useful. (P17, Research Chemist)

Collection driven digital hoarding led to purposely stored, well organized digital files. These participants had well labelled files and knew exactly what they had stored and how to access the files. If their work place had storage limits then they got around this by using external hard drives or cloud storage.

Everything is in my Dropbox, I have even got, so within Dropbox, I have got loads of folders for everything, for what I currently need to access, but I have also got two folders, which one is from my postdoc and one is from my PhD, as in the udrives from those Universities. So, it is basically my complete, I have got my whole, everything from my PhD period, everything from

my Postdoc all on Dropbox, in two folders, and then lots of folders for everything I have done since I have been here, basically. (P12, Associate Professor)

Where collection was an important motivation, there were few if any concerns about the amount of digital data that was being stored. This data was seen as valuable to participants now or would be valuable to them in the future. Here the collection dimension differs from the anxiety dimension in which data was kept just in case but regardless of its perceived value. A collection of digital data can provide a valuable resource for the owner and for others despite the passage of time or a change of circumstances as shown below.

I have got emails in my inbox from before my first spell of maternity leave and my daughter is nearly 8, so, it, I was in a different role then, but, I still get people coming back asking me questions, about those emails (P18, Senior Scientist)

Hoarding a valuable collection of digital data meant that these participants who display digital hoarding behaviours saw the stored digital data as part of their identity and felt a sense of pride as a result of having all this digital data. These participants were happy that other colleagues knew they kept a lot of data and were happy to respond to requests for information from their stored files.

I am often called upon by other members of staff, to say do you still have this, do you know this, because everyone knows I don't delete things. (P6, Senior Lecturer)

Deliberately choosing to keep all of their data meant that these participants who scored highly on the DHQ felt a high sense of ownership over their emails and their digital files and would certainly want to take their emails and digital files with them if they were ever to leave their job. Participants displayed a strong sense of ownership over their digital possessions and for some this feeling was long held.

Yes [Do you feel a sense of ownership], *and I have still for every single email I ever sent from my previous job, stored on a cloud. (P12, Associate Professor)*

While for others, the realisation that they felt a sense of ownership over their digital data was novel and almost a surprise to them.

I left and then someone else was like, oh yeah I will take your slides, like yeah that's fine, [sarcastically] like no! they are mine, I want to keep my slides they are really nice. (P11, PhD/Demonstrator)

Barriers to change

Most participants did not see any to change their current data management practices. Their workplaces had that assumed to be unlimited storage space and participants were happy to switch to use personal external drives or cloud storage in order to keep digital data stored.

> *Definitely, definitely, if there, if I am given limitless space then, if no one is limiting me, then I am going to store absolutely everything (P8, Senior Lecturer)*

Participants argued that keeping digital information was different from keeping physical data. Physical data, for example paper documents, could clutter an office and become a fire hazard. In the absence of these physical risks and associated safety concerns, participants struggled to articulate any real advantages to regular data deletion. They noted that search tools meant it was relatively easy to find digital files, even if they had thousands stored on their computer. In short, they felt that the amount of digital data stored did not have a meaningful impact on their everyday working practices.

> *The physical risk of fire having more paper around it takes that away, and you are not going to lose or destroy the data, or the files. (P9, Senior Research Assistant)*

Participants also noted that they received very little guidance on data deletion and retention. Most of the participants said they did not know their organizations retention and deletion policies, despite having received recent GDPR training. Most participants thought that receiving further guidance on what files should be retained or deleted would be beneficial.

> *But in terms of like digital files and deletion and stuff, I mean, we don't really get a lot of help with that.* (P16, PhD/Demonstrator)

Some participants were aware that their company hosted data management days, however, as completion is at the individual's discretion, many remained uncertain about which files should be retained or deleted. Despite the introduction of new policies and processes, staff often felt unsure about how to engage with new techniques for data management.

> *I think we have got, we have just introduced a new security tagging on files, with secret, highly restricted, restricted, so I think, in future, I haven't tried it yet, I don't, I don't know how to use it. (P20, Research Scientist).*

In terms of security, participants did not perceive any negative cyber security issues, believing that their company's security measures would be sufficient to keep their data secure. In addition, they felt that security issues were a concern for the organization rather for them as individuals. Where they did imagine some kind of data loss, some were dismissive, arguing that if everyone were affected then the consequences for any one individual would be more marginal:

And I do kind of think if there was some kind of data breach where my emails got hacked, so that people could get into my folder of applications then I'm sure it would be part of a larger data breach and I'm sure there would be more substantial people than me. (P12, Associate Professor)

The only security issue that was taken seriously, related to the physical security of the storage devices themselves and this was often in reference to personal devices. For example, participants discussed the implications of having their laptop or memory stick stolen, but ironically, thought that the solution to this problem was the duplication of data, i.e. ensuring backups existed on other systems. Overall, participants saw the benefits of keeping large amounts of digital data, and failed to recognise any downsides. If they did foresee any issue with security breaches, these were consequential for the company rather than for themselves.

DISCUSSION

The findings from our study extend the current, small body of knowledge around digital hoarding in particular the literature on workplace digital hoarding. We know that digital hoarding is an umbrella term capturing a range of different behaviors around the excessive accumulation of digital data and the difficulties deleting that data. Our study now provides a more nuanced way of conceptualising digital hoarding by considering the underlying motivations people have for hoarding their digital data in the workplace. Identifying the key motivations or dimensions of digital hoarding in this way is a useful first step in recognizing that individuals may respond differentially to tools and guidance aimed to reduce hoarding behaviours.

Our findings show that in this group of participants who scored highly on the DHQ, there were four underlying dimensions or motivations for their digital hoarding behaviour. We found that for some participants, ***anxiety*** drove digital hoarding in the workplace. This motivation draws on both a sense of keeping things, 'just in case', as noted before in relation to personal data contexts (Sweeten et al. 2018) but also a workplace specific anxiety around the need to be able to evidence work tasks and communications. Participants were fearful about deleting or losing data that may be needed in the future (Luxon et al., 2019) and this sense of anxiety may be one of the reasons for the gap between people's ideal deleting behaviour and their actual deleting behaviour (Alon & Nachmias, 2020). Keeping data in case it's needed in the future resonates with the more practical value of hoarding identified by Vitale et al., (2018) in relation to college students. The sense of anxiety around deleting data may be more acutely felt in knowledge intensive workplaces and for academics specifically where data may have both personal and organisational value. ***Disengagement*** was also a key driver for digital hoarding. This is a strong theme and one that builds upon the findings of Sweeten et al., (2018)

and Vitale et al., (2018) in both personal and workplace settings. The fact that disengagement was an important motivation for digital hoarding is perhaps more surprising in a workplace setting given that many information security policies specify data clean up behaviours for employees. While organizations' policies may encourage the deletion of unnecessary digital data, employees often feel this is not seen as a priority by managers. Participants feel reluctant to engage in an activity that is not assigned time within their workload and thus detracts from their key objectives (Beautement, Sasse, & Wonham, 2008). Where the policies are clearer, or the rules enforced then compliance was important.

Compliance, as a driver for digital hoarding, appears to be more specific to the workplace environment. Organizational procedures and policies may dictate a data management plan and so in effect drive the hoarding behaviour. This was more commonplace for participants working in team settings or on specific collaborative projects (Wiewiora, Trigunarsyah, Murphy & Coffey, 2013) and this may be a factor that is more sensitive to the organisational setting.

Collection, as a driver for digital hoarding has not been captured in relation to workplace hoarding before and is a novel finding from our study. Participants motivated by collection employed strategies similar to those used by structured filers (Henderson & Srinivasan, 2011) and used emails as prompts to action (see Gwizdka, 2004). Hoarding behaviours often pointed to individuals having a strong sense of ownership over their digital possessions (Cushing, 2011). Collection was an important driver for many especially around times of change or transition. Keeping digital data prevented knowledge from being lost (Holten et al., 2016) and created and strengthened identity roles for some of our participants.

The four dimensions compliment and extend our knowledge of digital hoarding behaviours and motivations. Understanding these motivations is key to the design and delivery of policies and tools that will support people to actively make decisions about the data they should be keeping and that which they should be deleting. Vitale and colleagues (Vitale 2019, Vitale, Odom & McGrenere, 2019) have successfully argued for a more personalised approach to the design of tools to support the deletion of personal data. We extend this argument and propose that tailored tools based on individual motivations as well as organisational context are also highly relevant to the discussion of data management tools and policies within a workplace setting. Below, we consider more closely the consequences of digital hoarding for the individual and for the organization taking into account the different underlying dimensions.

Consequences of digital hoarding

For the individual

Clearly, the consequences of digital hoarding to the individual are perceived as less significant than the consequences for the organization. Interestingly, where anxiety or collection were motivations for hoarding, individuals exhibited a strong sense of ownership around their digital data. This finding resonates with work by Cushing (2013) who noted that people differ in the extent that digital objects are seen as digital possessions – noting a tension between beliefs of possession and ownership. People may feel that digital objects are theirs but know that 'technically' they are owned by company or the organization. Academics in particular felt possessive over their digital data and talked about how they would take the data with them were they to change jobs. This is despite also recognizing that, in many cases, the data actually belonged to the organization.

For some of the participants, the accumulation of data symbolized their relative status and standing within the organization. Keeping finished work, copies of drafts and emails all documented the process of being in that role. Kaye et al., (2006) noted the importance of data management strategies in relation to the development and maintenance of being an academic. Building up a significant store of data in a physical space takes longer to do and may be more indicative of longevity. A physical store also gives visual cues to identity and highlights the academic as someone who authors papers, reads journals and assesses students work. This is a less visible practice in digital data but may well still provide a framework for building identity (Vertesi, Kaye, Jarosewski, Khovanskaya & Song, 2016). Across both organizations, hoarding behavior and the data itself allowed some participants (especially those motivated by compliance or collection) to present themselves in the role of 'information expert'. These people saw themselves as being particularly useful employees performing a vital 'go to role'. Those with a 'collection' of accumulated data exhibited pleasure in having a certain sense of control over the data and recognizing who needed what data and when. For example, these participants might keep a lot of data that makes sense to them, but would extract from this store smaller, more meaningful pieces of data if they were sharing it with a colleague.

In this study, we identified collection as a key dimension of hoarding behavior. In fact, in the physical hoarding literature the arguments around this distinction are well rehearsed. Belk et al. (1988) describes how the two terms are distinguishable based on how the items or possessions relate to the construction of self. Collections, unlike hoards, have specific meanings and structures. To our knowledge, this is the first time this distinction has been identified in digital hoarding and collecting. The way in which our 'collectors' and to a certain extent those participants whose behavior was driven by compliance described their digital data practices resembled the behaviors of collectors as opposed to hoarders. Within a workplace setting, it is also important to understand the extent to which hoarding behaviours are motivated not only by compliance (to data protection and workplace requirements) but also by disengagement, anxiety and collection. The extent to which people were

comfortable with their hoarding identity, the label and even the term itself underlies issues with class and status that have been likened to the way in which physical hoarders and collectors are perceived (Nordsletten et al., 2013).

For the organization

The prevailing organizational culture within the workplace affected the type of hoarding behaviors and motivations reported. Compliance was particularly strong for participants working in close project teams where there was a strong sense of collaboration between colleagues. Data was being 'hoarded' but this behavior was seen as being for the good of the group rather than for the individual's benefit per se. Organizations that encourage participation, teamwork and informality are more likely to engender this kind of sharing, whereas those that value individual power and competition amongst employees are more likely to lead to hoarding as a personal and powerful act of collecting valuable information that can be used to define their role within the organization as well as demonstrate compliance (Wiewiora et al., 2013). Of course, 'compliant' hoarding does not necessarily lead to individuals withholding information. The insecurities of the employees is another factor likely to impact upon the levels of hoarding (Gormley & Gormley 2012). Where anxiety was a key dimension of the hoarding, individuals were driven to keep data 'just in case' or hold it as useful evidence for a future problem or dispute, a finding that resonates with the work by Sweeten et al., (2018).

Digital hoarding can be used to generate feelings of superiority where knowledge is kept from others deliberately (Gormley & Gormley 2012). Information hoarders are not always information sharers – and can choose to keep information away from others in strategic ways (Evans et al., 2014). Interestingly, in this study, we encountered only a few occasions where participants were reluctant to share and this may be a feature of the individual nature of the work they were describing. While other studies have noted the negative perceptions of physical hoarding (Oravec, 2015), for many of our participants there was a certain kudos associated with being a 'digital hoarder' and the behavior was not seen as overly problematic or stigmatized.

Design implications for digital hoarding

Organizations may not realize the extent of digital data that individuals accumulate. Much of the accumulated data is likely to contain personal information. For example, personal information about candidates for a post or personal data around student circumstances may be kept as undeleted emails or as saved documents, either intentionally or accidentally. Ultimately, such practices can render both the individual and the organization at risk of non-compliance with GDPR, particularly pertinent given the fact that the majority of our participants identified as having a data protection responsibility. Participants driven by a compliance motivation are likely to pose less of a threat in this context, as

they only keep data that they have been instructed to keep and are happy to delete data once it is no longer needed. However, disengagement and anxiety motivate individuals to hoard large volumes of unorganized data. These people are not often aware of all the data they have saved, and so are at risk of having highly personal data unintentionally saved on their personal computer or their organization's cloud storage. It is therefore important that employees are made aware of the way their own hoarding behaviors might implicate themselves and their organization in terms of GDPR breaches.

Our participants were usually unaware of any policies or guidelines regarding data management (with the exception of material subject to ethical considerations) and there was often a lack of clear policy within the organization concerning data management more generally. Most participants mentioned that to get around storage limits, they used cloud platforms. These are popular choices for individuals because they offer large or unlimited storage space that requires little maintenance from the user. However, cloud platforms can raise additional privacy concerns that users may not fully understand (Odom, Sellen, Harper & Thereska, 2012). Organizations should be aware if their employees are using cloud platforms to store data and what data they are storing on these platforms, as well as if this data is stored securely. More guidance on retention and deletion policies when using cloud platforms for work data may be useful for ensuring that employees are following GDPR guidelines when storing data on cloud platforms.

That said, we already know from the wider literature on usable cybersecurity, that (a) policy compliance is a huge problem (Bulgurcu, Cavusoglu & Benbasat , 2010; Herath & Rao, 2009) in part because (b) employees are focused upon the 'day job' and generally do not see data management as a priority (Beautement, Becker, Parkin, Krol & Sasse, 2016; Kirlappos, Beautement & Sasse, 2013). On this last point, our participants noted that the time resources needed to manage data more efficiently and effectively were simply not factored into their working day. Participants believed that their managers would be unhappy to find them spending time tidying up files or emails. Of course, organizations could take steps to reduce their mass emails and consider more carefully who exactly needs to see the information. Or they could engage in stronger measures such as bringing in 'deletion' policies, whereby emails are routinely deleted after a certain period of time. Oravec (2015) even describes the possibility of organizations updating or even deleting data on users' devices without their explicit consent. Such measures are unlikely to be popular and more work needs to be done to ensure that employees engage with best practice around data management rather than focusing on compliance alone.

A more user-centred approach would be to focus on empowering users to take more control over their own data. Vitale et al., (2018) identified a number of tools designed to make it easier for users to be aware of and think more carefully about the data they have stored. Other tools identify how and where users are storing data. They provide information including the largest files stored, when they were last accessed and can also provide the user with suggestions as to how to free up storage space on devices. Such features and applications may be useful in providing employees with greater self-awareness although as yet there is no data on how these tools are really being used. Vitale et al., 2019 have suggested that tools need to match individuals' needs and preferences with regards to personal data management. On the basis of the work we've presented here, we'd also suggest that it would be fruitful to design resources that support the different dimensions of hoarding. Two dimensions in particular might present problems for the organization: disengagement and anxiety. Designing for 'disengagement' might demand very little of the end user and might involve the use of default settings whereby data is automatically classified with sensitive material flagged for deletion. Designing for 'anxiety' on the other hand would involve prompting the user to tag emails and data files more thoughtfully, with some kind of classification that supports their particular concerns.

Limitations and future work

The focus of this study has been on digital hoarding within large organizations with a largely academic sample. It may be that in smaller organizations or in job settings not related to academia or research more broadly the values and culture are different and that different issues regarding digital hoarding are prevalent. Whilst some concerns will remain constant, the larger volume of data passing through larger organizations means that they are of particular interest in this regard. Different organisations including more generic office-based workplaces may exhibit different practices and attitudes towards digital data hoarding and this is something that we are currently exploring. Our current focus on academic and scientific organisations provides practical implications of digital hoarding for these settings but it remains to be seen whether these findings are more generalizable across a diverse range of workplace contexts.

In future work it would be interesting to develop interventions that take account of the dimensions of hoarding, or to see if these dimensions are useful in identifying crucial time points at which a relevant intervention would be useful. For example, disengagement seems to drive hoarding behavior over a longer period of time. Here, individuals can only accumulate large amounts of data if they have been with the organization for a longer period. Those new to the organization will have had less time to accumulate data and feel less ownership over that data. We are also currently considering the role of age in relation to digital hoarding. We aim to assess whether those individuals who have grown up with digital storage and instant access to digital information resources may not be as likely to engage

in hoarding behavior if they believe they can access the resources as and when they need from centrally held repositories (Gormley & Gormley, 2012).

Conclusion

This study examined the digital hoarding behaviors of employees working in two large research based organizations. Although all participants had scored highly on the DHQ, we identified four underlying dimensions of their digital hoarding practices (anxiety, disengagement, compliance and collection). Going forward the potential impact of strategies to support digital decluttering will need to encompass these different digital hoarding motivations across a range of organizational settings. Different tools may be effective within certain workplace cultures and providing people with an awareness of their current hoarding behavior or indications as to how to start decluttering may be important. Understanding how and why different users would engage with such tools warrants further exploration.

Acknowledgement

This research forms part of a project relating to the cybersecurity risks of digital hoarding which was funded by the Centre for Research and Evidence on Security Threats—an independent Centre commissioned by the UK Economic and Social Research Council (ESRC).

References

Alon, L., & Nachmias, R. (2020). Gaps between actual and ideal personal information management behavior. *Computers in Human Behavior*, *107*, 106292.

Beautement, A., Becker, I., Parkin, S., Krol, K., & Sasse, M. A. (2016, June). Productive security: A scalable methodology for analysing employee security behaviours. In *12th Symposium on Usable Privacy and Security (SOUPS)* (pp. 253-270).

Beautement, A., Sasse, M. A., & Wonham, M. (2008, September). The compliance budget: managing security behaviour in organisations. In *Proceedings of the 2008 New Security Paradigms Workshop* (pp. 47-58).Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS quarterly*, *34*(3), 523-548.

Belk, R.W. M. Wallendorf, & J. Sherry, (1988). Collectors and collecting. *ACR North American Advances*.

Braun, V. & V., Clarke, (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

CGOC Information Governance Process Maturity Model, January, 2017. Accessed, 25th January, 2017 from: http://www.cgoc.com/

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.

Cushing, A L. (2013). "It's stuff that speaks to me": Exploring the characteristics of digital possessions. Journal of the American Society for Information Science and Technology, 64(8), 1723-1734.

Evans, J. M. Hendron, M.G & J.B. Oldroyd. (2014). Withholding the Ace: The Individual- and Unit-Level Performance Effects of Self-Reported and Perceived Knowledge Hoarding. *Organization Science* 26.2: 494-510.

Gerritsen, D. B. D. Tasse, J. K. Olsen, T. A. Vlahovic, R. Gulotta, W. Odom, & J. Zimmerman. (2016, May). Mailing archived emails as postcards: Probing the value of virtual collections. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 1187-1199). ACM.

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

Gulotta, R. W. Odom, J. Forlizzi, & H. Faste, (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). ACM.

Gwizdka, J. (2004, April). Email task management styles: the cleaners and the keepers. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1235-1238). ACM.

Henderson, S., & Srinivasan, A. (2011, January). Filing, piling & structuring: strategies for personal document management. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on* (pp. 1-10). IEEE.

Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of Information Systems*, *18*(2), 106-125.

Holten, A.L. G.R Hancock, R Persson, A.M Hansen, & Hogh, A. (2016). Knowledge hoarding. Journal of Knowledge Management, 20(2), 215-229.

Information Commissioners Office (2018). Guide to General Data Protection Regulation. https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation

Kirlappos, I., Beautement, A., & Sasse, M. A. (2013, April). "Comply or Die" Is Dead: Long live security-aware principal agents. In *International Conference on Financial Cryptography and Data Security* (pp. 70-82). Springer, Berlin, Heidelberg.

Kaye, J, J, K Vertesi, , S. David, L. Onaga, & T. Pinch, (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). ACM.

Kibby, M. (2009). Collect yourself: Negotiating personal music archives. *Information, Communication & Society, 12*(3), 428-443.

Luxon, A. M., Hamilton, C. E., Bates, S., & Chasson, G. S. (2019). Pinning our possessions: Associations between digital hoarding and symptoms of hoarding disorder. *Journal of Obsessive-Compulsive and Related Disorders*, 21, 60-68.

Massey, C. S. TenBrook, C. Tatum, & S. Whittaker. (2014). PIM and personality: what do our personal file systems say about us? *Proceedings of the 32nd annual ACM conference on Human Factors in Computing Systems*: 3695-3704.

Neave, N. P. Briggs, K. McKellar, & E. Sillence. (2019). Digital hoarding behaviors: measurement and evaluation. *Computers in Human Behavior,* 96, 72-77.

Nordsletten, A,E., Reichenberg, A., Hatch, S.L., Fernández de la Cruz, L., Pertusa, A., Hotopf, M., and Mataix-Cols, D. (2013). Epidemiology of hoarding disorder. *British Journal of Psychiatry*, 203: 445-452.

Odom, W. A, Sellen, R, Harper, and Eno Thereska. (2012). Lost in Translation: Understanding the Possession of Digital Things in the Cloud. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 781–790. DOI: http://dx.doi.org/10.1145/2207676.2207789

Odom, W., Zimmerman, J., & Forlizzi, J. (2011, May). Teenagers and their virtual possessions: design opportunities and issues. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 1491-1500).

Oravec. J. A. (2015). Depraved, Distracted, Disabled, or Just "Pack Rats"? Workplace Hoarding Persona in Physical and Virtual Realms. Persona Studies, 1(2), 75-87.

Oravec. J. A. (2017). Digital (or Virtual) Hoarding: Emerging Implications of Digital Hoarding for Computing, Psychology, and Organization Science. International Journal of Computers in Clinical Practice (IJCCP), 3(1), 27-39.

Oravec. J. A. (2018). Virtual Hoarding. In Encyclopedia of Information Science and Technology, Fourth Edition (pp. 4306-4314). IGI Global.

Sedera, D., & Lokuge, S. (2018). Is Digital Hoarding a Mental Disorder? Development of a Construct for Digital Hoarding for Future IS Research. In ECIS 2018 (pp. 1-1). Association for Information Systems.

State of California Department of Justice (2020). California Consumer Privacy Act (CCPA) https://oag.ca.gov/privacy/ccpa)

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

Thomas, L. P. & Briggs, (2014, October). An older adult perspective on digital legacy. In Proceedings of the 8th nordic conference on human-computer Interaction: Fun, fast, foundational (pp. 237-246). ACM.

Van Gennip, D. E. Van Den Hoven, & P. Markopoulos, (2015, April). Things that make us reminisce: Everyday memory cues as opportunities for interaction design. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3443-3452). ACM.

Van Bennekom, M. J., Blom, R. M., Vulink, N., & Denys, D. (2015). Case Report: A case of digital hoarding. *BMJ case reports*, *2015*.

Vitale, F. (2019, May). Designing for Long-term Digital Data Management. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-4).

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).

Vitale, F. I. Janzen, & J. McGrenere, (2018, April). Hoarding and Minimalism: Tendencies in Digital Data Preservation. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 587). ACM.

Vertesi, J. J. Kaye, S. N. Jarosewski, V.D. Khovanskaya, & J. Song, (2016, February). Data Narratives: Uncovering tensions in personal data management. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 478-490). ACM.

Watkins, R.D. A. Sellen, & S.E. Lindley. (2015, April). Digital collections and digital collecting practices. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3423-3432). ACM.

Whittaker, W & C. Sidner, (1996). Email overload: exploring personal information management of email. Proc. CHI, ACM Press 276-283

Whittaker. S (2011). Personal information management: from information consumption to curation. *Annual review of information science and technology*, *45*(1), 1-62

Wiewiora, W B. Trigunarsyah, G. Murphy, & V. Coffey. (2013). Organizational culture and willingness to share knowledge: A competing values perspective in Australian context. International Journal of Project Management, 31(8), 1163-1174