



# Cybersecurity Risks of Digital Hoarding Behaviours

**FULL REPORT** 

AUGUST 2020

Prof Nick Neave Prof Pam Briggs Dr Elizabeth Sillence Dr Kerry McKellar

## **Cybersecurity Risks of Digital Hoarding Behaviours**

## **Full Report**

**Prof Nick Neave Prof Pam Briggs** Dr Elizabeth Sillence Dr Kerry McKellar

This report is from Cybersecurity Risks of Digital Hoarding Behaviours - a commissioned project funded by CREST.

The project explores the psychological characteristics of individuals who engage in digital hoarding, and the risks these behaviours can pose to organisations.

To find out more information about this project, and to see other outputs from the team, visit the CREST website at: www.crestresearch.ac.uk/projects/cvber-security-risks-digital-hoarding/

#### About CREST

The Centre for Research and Evidence on Security Threats (CREST) is a national hub for understanding, countering and mitigating security threats. It is an independent centre, commissioned by the Economic and Social Research Council (ESRC) and funded in part by the UK security and intelligence agencies (ESRC Award: ES/N009614/1). www.crestresearch.ac.uk





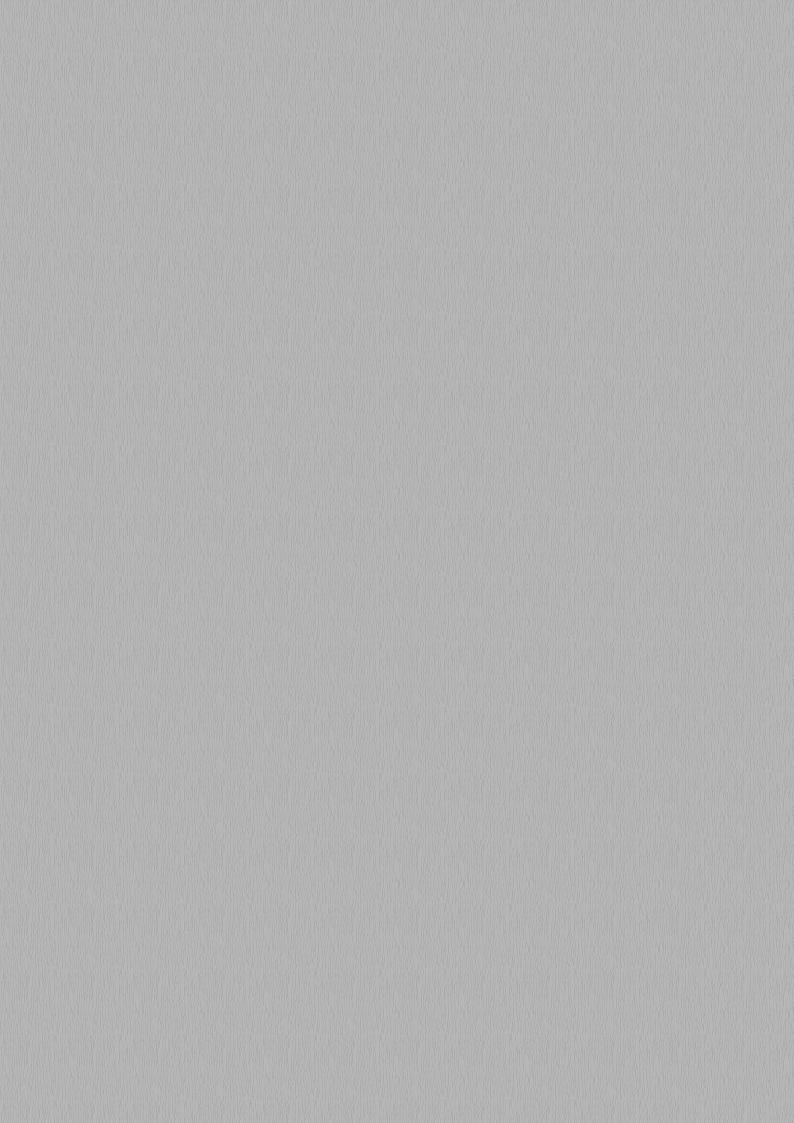






## TABLE OF CONTENTS

EXECUTIVE SUMMARY	5
BACKGROUND	6
ADVICE FOR ORGANISATIONS	7
KNOWLEDGE GAPS	10
STUDY 1	11
STUDY 2	13
STUDY 3	18
DIMENSIONS OF DIGITAL HOARDING IN THE WORKPLACE	20
1. Disengagement	20
2. Compliance	21
3. Anxiety	22
4. Collection	23
REFERENCES	26
APPENDIX: DIGITAL BEHAVIOURS QUESTIONNAIRE (DBQ)	28





## EXECUTIVE SUMMARY

Digital hoarding has been defined as "...the accumulation of digital files to the point of loss of perspective, which eventually results in stress and disorganisation" (van Bennekom et al., 2015). Whilst there is clearly no adverse impact on physical living spaces, personal and professional life may still be negatively affected by such behaviours and the host organisation may suffer as a function of operational inefficiencies resulting from excessive digital clutter (Gormley and Gormley, 2012).

In addition, digital hoarding has clear implications for cybersecurity and data protection within organisations. With this in mind, developing a greater understanding of the different characteristics of individuals with data hoarding tendencies will enable organisations to mitigate these cybersecurity risks.

The aim of this research project was to develop a new measure for the measurement of digital hoarding and assess the extent to which digital and physical hoarding were related.

Through three studies we developed and validated a new questionnaire to measure digital hoarding behaviours in the workplace. We found that:

- In a range of organisations, digital hoarding is very common (especially for emails).
- Individuals with data protection responsibilities accumulate and retain more digital files than those without such responsibilities.
- Lack of deletion is typically driven by the demands of the job.
- Individuals are aware that digital hoarding might lead to negative consequences for themselves and their organisation yet persist in this practice.
- Individuals scoring higher in physical hoarding personality characteristics are more likely to display digital hoarding behaviours.
- There appears to be four different types of digital hoarders: the 'Collector', the 'Accidental Hoarder', the 'Hoarder by Instruction' and the 'Anxious Hoarder'.









### **BACKGROUND**

Recently, there has been much speculation in the media and in online forums concerning the existence of and potential problems relating to digital hoarding. The existence of digital hoarding is not surprising given that our lives are becoming increasingly digital (Gulotta et al., 2013) and that contemporary presentations of the self extend to digital possessions (Cushing, 2011, 2013). Yet relatively little research describes the accumulation of digital possessions and addresses the potentially unique issues and problems that digital hoarding might create.

Digital hoarding has been defined as "...the accumulation of digital files to the point of loss of perspective, which eventually results in stress and disorganisation" (van Bennekom et al., 2015). Whilst there is clearly no adverse impact on physical living spaces, personal and professional life may still be negatively affected by such behaviours and the host organisation may suffer as a function of operational inefficiencies resulting from excessive digital clutter (Gormley and Gormley, 2012). Whilst there are very few investigations of digital hoarding, two studies are worth reporting here. Firstly, van Bennekom et al. (2015) report the case of a physical hoarder who then became obsessed with the hoarding of digital photographs. He displayed high levels of attachment to the digital images, which he couldn't bear to discard, and the time he spent organising the many thousands of images interfered with his daily functioning and caused him great distress - classic symptoms of physical hoarding.

Secondly, Vitale et al. (2018) conducted a qualitative investigation of both hoarding and minimalist data storing practices. Their work involved 23 participants who were asked not only to talk about the value of their digital data and relive their data 'life history' but also to demonstrate their data storage practices using their own devices. In their study, digital hoarding had both an emotional and a practical component for people, but the authors noted that hoarding existed on a 'spectrum' and argued that more research was needed to understand this spectrum.

Personal Information Management (PIM) is the more general term used to describe how individuals collect, store, organise and retrieve digital items (Boardman and Sasse, 2004). PIM is time-consuming and burdensome (Lansdale, 1988) and analyses of email deletion and archiving behaviours in organisations show that users do not manage digital information in an effective way. They typically keep half of the emails they receive and reply to about a third of them (e.g. Dabbish et al., 2005), with very few people engaging in proactive 'clean-up' of that stored information (Bergman and Beyth-Marom, 2003). Data clutter is thus highly prevalent, and the impact of data hoarding on a business can be large as cost, data lifespan, effectiveness, productivity and knowledge management can all be adversely affected by excessive data hoarding (Gormley and Gormley, 2012). 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).

Digital hoarding could clearly have negative consequences for the individual and for organisations, but information relating to digital hoarding behaviours remains sparse, especially in relation to the underlying motivations. Sweeten et al. (2018) conducted a qualitative assessment of digital hoarding behaviours, motivations and consequences in 45 individuals and identified themes common to physical hoarding, such as the over-accumulation of digital materials, difficulties in deleting such materials, and feelings of anxiety relating to this accumulation/difficulty deleting. The authors noted that the ability to properly identify and quantify digital hoarding behaviours was currently lacking, and so the aim of the first study was to develop and validate a psychometric questionnaire to identify the extent of digital hoarding behaviours in the workplace and to explore the extent to which digital hoarding was associated with the known characteristics of physical hoarding digital hoarding.

The aim of this research project was to develop a new measure for the measurement of digital hoarding and assess the extent to which digital and physical hording were related.

## ADVICE FOR ORGANISATIONS

Digital hoarding is commonplace (especially for emails) and appears particularly prevalent in individuals with data protection responsibilities. It is thus partly driven by the demands of the job, as individuals are hoarding data that is imposed upon them with good intentions, even though they realise that such behaviours might have negative consequences for themselves and their organisation.

Our hoarding questionnaire could enable organisations to gain a quantitative understanding of the amount and type of files that employees are routinely keeping, and to perhaps explore subgroups within the organisation to see if such practices are being created and enforced by specific policies (e.g. email deletion polices, data retention policies etc.) that might generate inadvertent hoarding behaviours within certain elements of the workforce. This appears to be particularly true for employees with data protection responsibilities, as they retain significantly more information than individuals without such responsibilities.

Indeed, there is a strong relationship between the physical hoarding characteristics of accumulation and difficulty discarding and digital hoarding. 'Accumulators' seem reluctant to delete emails, whilst those high in 'difficulty deleting' are able to delete emails but seem more reluctant to delete other more substantial files (text files, spreadsheets, presentations, photographs). This makes sense when we consider that the 'difficulty deleting' factor represents a more emotional attachment to files and presumably operates on digital items with more personal value – i.e. files that represent something that the participant has created or contributed. The 'accumulating' factor reflects a tendency to keep all information that is shared on the grounds that it may be useful at some point and it would make sense that those who score highly in 'accumulating' are loath to delete emails. It is possible that individuals with hoarding propensities are being drawn to certain job roles which entail the accumulation and storage of data, but at the moment this remains speculative.

In one particular regard, employee hoarding behaviour is likely to become very troublesome. The roll-out of new privacy and data protection legislation that regulates the storage of personal data (e.g. GDPR in Europe) can mean that both organisations and individuals could be unwittingly storing data illegally. Oravec (2015) discusses the ways that employees can be blind to legal issues of data handling and data protection in the workplace, rendering the organisation liable to legal action with possible financial penalties. She also notes, in line with Sprague (1983), that aggressive managerial solutions to such problems are likely to backfire as they typically involve increased surveillance resulting in workplace privacy concerns. Identifying the key motivations or dimensions of digital hoarding in this way is a useful first step in recognising that individuals may respond differentially to tools and guidance aimed to reduce hoarding behaviours.

Interestingly, where anxiety or collection were motivations for hoarding, individuals exhibited a strong sense of ownership around their digital data. This finding resonates with the work of Cushing (2013) who noted that people differ to the extent that digital objects are seen as digital possessions - noting a tension between beliefs of possession and ownership in her study. People may feel that digital objects are theirs but know that 'technically' they are owned by the company or organisation. 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 recognising that, in many cases, the data actually belonged to the organisation. Across both organisations hoarding behaviour and the data itself allowed some hoarders (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 recognising who needed what data and when. For example, these hoarders 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.

#### **ADVICE FOR ORGANISATIONS**

Cybersecurity risks of digital hoarding behaviours

In this study, we identified collection as a key dimension of hoarding behaviour. 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 our hoarders, by instruction described their digital data practices resembled the behaviours of collectors as opposed to hoarders. Given a definition (based on a single case) it's perhaps unsurprising that we are beginning to see the need for a more nuanced approach to digital hoarding behaviours. 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).

The prevailing organisational culture within the workplace affected the type of hoarding behaviours and motivations reported. Compliance was particularly strong for hoarders working in close project teams where there was a strong sense of collaboration between colleagues. Data was being 'hoarded' but this behaviour was seen as being for the good of the group rather than for the individual's benefit per se.

Organisations that encourage participation, teamwork and informality are more likely to engender this kind of sharing, whereas those that value individual power and completion 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 organisation 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 employee is another factor likely to impact upon the levels of hoarding. Where anxiety was a key dimension of the hoarding, hoarders 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 and 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 digital hoarders were reluctant to share, and this may be a feature of the individual nature of the work they were describing. Whilst 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 behaviour was not seen as overly problematic or stigmatised.

Organisations may not realise the extent of digital data that individuals accumulate; much of it likely to contain personal data. 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 organisation at risk of non-compliance with GDPR. Hoarders 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 unorganised 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 organisation's cloud storage. It is therefore important that employees are made aware of the way their own hoarding behaviours might implicate themselves and their organisation 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 organisation concerning data management more generally (Oravec, 2017). 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 (Adom et al., 2012). Organisations should be aware if their employees are

#### **ADVICE FOR ORGANISATIONS**

Neave, Briggs, Sillence, McKellai

using cloud platforms to store data, and what data they are storing on these platforms, and whether 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 et al., 2010; Herath and Rao, 2009) in part because (b) employees are focused on the 'day job' and generally do not see data management as a priority (Beautement et al., 2016; Kirlappos et al., 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, organisations 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 (2018) even describes the possibility of organisations 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 on how to free up storage space on devices. Such features and applications may be useful in providing hoarders with greater self-awareness, although as yet there is no data on how these tools are really being used.

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 organisation: 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.

## KNOWLEDGE GAPS

- 1. One issue is that we simply don't yet understand the scale of the issue. Further work is needed to understand the types of digital hoarding behaviours in the workplace. We have begun this with our discussion of those who find it distressing to delete files and those who are more intent on accumulating, and in the analysis of individuals with and without data protection responsibilities, but as yet we don't understand enough about the kinds of material these individuals hoard and the associated workplace risks.
- 2. The samples did not specify particular types of employees, job roles or types of organisations, and such variables may have a large impact upon the behaviours observed. Different organisations and job roles may impact upon the opportunity for digital hoarding, for example some organisations have mandatory email deletion policies, and so the questionnaire may not be applicable or useful in certain occupational settings.
- 3. It must be noted that differentiating between those with and without data protection responsibilities was accomplished by a single question: "Does your current job role entail any responsibility for data protection?" It is possible that this question may lack sufficient clarity to enable some individuals to make an accurate assessment, and this issue clearly requires additional research. In Europe, now that General Data Protection Regulations (GDPR) have come into force, many more employees have data protection responsibilities, and employees are coming to realise that they have data protection responsibilities at a personal level but may not fully understand the implications of this.
- 4. Digital hoarding behaviours could of course have an impact outside of the workplace, many people are 'freelance' or work from home, and so future work could consider these additional factors in more detail, though it would be simple to adapt the questionnaires to address personal, rather than workplace behaviours.

- 5. An additional knowledge gap relates to our focus on emails (justified as such files were the most commonly hoarded type of file), but other types of files could be more problematic in different organisations. Within people's personal lives it is also likely that emails will figure strongly, but we might also expect to see the hoarding of music files, photographs, apps and social media 'friends' for example.
- 6. We have identified different types of digital hoarder from focus group interviews, but we now need to see if these 'types' hold up to quantitative scrutiny and can be measured by the development of a new questionnaire.
- 7. The focus of this set of studies has been on digital hoarding within large organisations. It may be that in smaller organisations 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 organisations means that they are of particular interest in this regard.
- 8. A current gap in knowledge relates to the development of training programmes or interventions that take account of the dimensions of hoarding we have found, 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 behaviour over a longer period of time. Here, hoarders can only accumulate large amounts of data if they have been with the organisation for a longer period. Individuals new to the organisation will have had less time to accumulate data and feel less ownership over that data.
- 9. 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 behaviour if they believe they can access the resources as and when they need from centrally held repositories (Gerritsen et al., 2016).

#### **SAMPLE**

424 individuals took part in this phase, comprising 208 males (aged 21–58, mean = 52.9, SD = 13.4) and 216 females (aged 20–75, mean = 45.2, SD = 12.6). 268 participants were in full-time employment and 137 in part-time employment (19 participants did not complete this question); of the 268 responders, 123 (60.9%) had data protection responsibilities. Participants were obtained through a market research agency (Critical Mix – www.criticalmix.co.uk), which was asked to pinpoint individuals who were currently in work and used a computer as part of their job. All participants were over 18 and resided in the UK.

#### **MATERIALS**

The initial questionnaire (the Digital Hoarding Questionnaire: DHQ) contained 12 statements adapted from the physical hoarding literature which focussed on the core facets of accumulation/clutter, difficulty discarding, and distress (Frost and Gross, 1993; Steketee and Frost, 2003) with sample items including 'I find it extremely difficult to delete old or unused files' and 'I tend to accumulate digital files, even when they are not directly relevant to my job'. Items were scored on a Likert-type scale from 0 (not at all) to 5 (very much so).

In addition, a second questionnaire (the Digital Behaviours in the Workplace Questionnaire: DBWQ) was created to assess the extent of digital hoarding in the workplace, asking about digital files stored, deletion behaviours, and beliefs about the consequences of digital hoarding to the self and the organisation. The questionnaire opened with nine questions on demographics and information relating to employer size, length of time with employer, length of time in job role, and whether or not the individual held data protection responsibilities. Subsequent questions asked about the number of digital files of various types possessed, how often such files were deleted, reasons for not deleting emails, and the potential personal and professional consequences of not deleting.

These questions were derived from the existing literature relating to digital hoarding behaviours and informal discussions within the research team. Of particular importance were the findings from Sweeten et al. (2018) whose '5 barriers to deletion' (p56) are included in section 3 of our questionnaire, and the problems they identified with accumulating digital data (p58) influenced section 4, where we ask about the potential consequences of digital hoarding. Both sections can be combined to form a Digital Behaviours Questionnaire (DBQ); see Appendix for final version.

#### **PROCEDURE**

Following ethical approval, a link to the questionnaires on the online survey platform Qualtrics (www. qualtrics.com) was created. Participants fulfilling the recruitment criteria (adults, working with computers etc.) were recruited by the market research agency. On being directed to the surveys, participants gave their informed consent, provided information about their age and gender, and were then asked to complete both the DHQ and the DHWQ. On completion they were directed to a debrief page.

#### **RESULTS**

This pilot study generated a wealth of data which will not be covered in detail as the key aim was to refine (where necessary) the two questionnaires. Key facts to note were that participants generated a wide range of scores indicating that the questions were able to differentiate individuals who hoard digital items from those who do not. We did not compare employees with and without data protection responsibilities at this stage. We conducted a principal component analysis with varimax rotation. Both KMO and Bartlett's test indicated that the sampling was sufficient (KMO = . 914,  $\chi 2$  (66) =2656, p < .001 ). The scree plot again suggested a two-factor solution, with the two factors accounting for 62.9% of the variance. Based on the highest loadings for each factor, this resulted in two scales, one comprising six items (Cronbach's alpha = .905) which we call 'difficulty deleting' and the other comprising four items (Cronbach's alpha = .824) which we call 'accumulating'. Both scales correlated substantially (r = .736). Difficulty deleting sub-scale scores ranged from 0-29 (Mean=8.3, SD=7.4), the accumulating sub-scale scores ranged from 0-24

Cybersecurity risks of digital hoarding behaviours

(Mean=10.3, SD=6.2). Two items did not fit with either of the factors and so were removed; this resulted in a 10-item scale. Items relating to emotional discomfort were found to be associated with the factor of 'difficulty deleting'. These two factors were strongly positively correlated with one another, and to hoarding behaviours at work as measured by the DBWQ, specifically with the total number of digital files retained, and difficulty in deleting certain files.

This first study thus generated the data necessary to devise more robust questionnaires, better suited to the assessment of digital hoarding attitudes and behaviours. The end result was a Digital Hoarding Questionnaire (DHQ) designed as a psychometric assessment of digital hoarding traits and attitudes; and a Digital Behaviours at Work Questionnaire (DBWQ) which included individual and workplace demographics (nine items) and four sections on workplace hoarding behaviours and attitudes that measured (i) accumulation and storage behaviours; (ii) deletion behaviours; (iii) rationale for keeping emails and (iv) perceived consequences for self and company. These are described in more detail in Study 2, below.

#### **SAMPLE**

203 individuals took part in this phase, comprising 97 males (aged 25–78, mean = 48.8, SD = 11.2) and 105 females (aged 20–79, mean = 46.6, SD = 12.3). All of the participants were currently working and regularly used computers as part of their job. 147 participants were in full-time employment and 56 in part-time employment; of these 68 had data protection responsibilities.

#### **MATERIALS**

The final versions of both questionnaires comprise, firstly, the revised 10-item DHQ, with items that related to the earlier two-factor structure: difficulty deleting and accumulation. The DHQ was scored from 0 (not at all) -7 (very much so), difficulty deleting scores ranged from 0–42 (Mean=16.8, SD=10.1) accumulation scored ranged from 0–35 (Mean=19.0, SD=7.9). Secondly, the DBWQ, with nine individual and workplace demographic items followed by four sections: two capturing the hoarding behaviours and types of hoarded items, and two addressing the reasons for those behaviours and their implications for the individual and the work organisation.

Section 1 focussed on accumulation and storage behaviours, providing a list of common digital files, and asking the respondent to use a slider (ranging from 0 to 1000+) to identify how many files of each type they have right now (this was completed electronically and gave a precise value, but could easily be adapted for pencil-and-paper by asking people to state how many files they currently have).

Section 2 focussed on deletion behaviours, asking about deletion activity for the same list of files from section 1, with a choice of five options ranging from 'I typically delete these daily' to 'I hardly ever delete these files'. With each box allocated a numerical value of 1–5 (daily to hardly ever) then a total score of 40 is possible; a higher score indicating greater hoarding behaviour (less deleting activity).

Section 3 focussed specifically on emails, and the reasons why individuals do not delete emails. A list of 12 possible reasons why individuals do not delete emails was provided (e.g. 'I don't delete them because I simply don't have time to delete them all') and respondents were asked to indicate for each reason how true that was for them on a seven-point Likert-type scale (1 = not at all, 7 = very much so). As the aim here was to identify a common list of reasons for not deleting emails, a score is not generated.

The final section addressed an individual's perceptions of the personal and organisational consequences if stored files were made public or stolen. The same list of file types were presented as per sections 1 and 2, and respondents were asked to identify the consequences to them personally and then the consequences for their company if such files were to be made public on a seven-point Likert-type scale (1 = no consequence at all, to 7 = very severe consequences). As the aim here was to identify perceptions relating to the consequences of files being released, a total score is not generated.

#### **PROCEDURE**

The procedure followed that outlined in study 1, with the exception of the addition of a test-retest assessment. Here, a random sample of 50 individuals were asked to retake the study again six weeks after first taking part so that we could establish the test-retest reliability of the scale. All scales were significantly correlated from time one to six weeks later during the test-retest (provided in the results section). In section 3, rationale for keeping emails question, item-to-item correlations were completed as these are individual items that cannot be totalled together; all items significantly correlated during the test-retest apart from one item in question 'I don't delete something in case I need evidence that something has been done' (r=.234, p=102).

#### **RESULTS**

Note that the sample size varies somewhat across analyses because some respondents did not answer all items.

## THE DIGITAL HOARDING QUESTIONNAIRE (DHQ)

The DHQ contained 10 items associated with the propensity to hoard digital items. In order to check if the data showed the same two-factor structure identified in

Study 1, a principal components analysis with varimax rotation was conducted. Both KMO and Bartlett's test indicated that the sampling was sufficient (KMO= . 927,  $\chi 2$  (45) =1762, p < .001). The scree plot again suggested a two-factor solution, with the two factors accounting for 77.92% of the variance. The factor loadings are presented in Table 1. Based on the highest loadings for each factor, this resulted in two scales, one comprising six items (Cronbach's alpha = .945) which we call 'difficulty deleting' and the other comprising four items (Cronbach's alpha = .873) which we call 'accumulating'. Both scales correlated substantially (r = .756)

	Factor 1: Difficulty deleting	Difficulty deleting item-total correlations	Factor 2: Accumulating	Accumulating item-total correlations
1 I find it extremely difficult to delete old or unused files	.717	r=.822, p<.001	.387	-
2. I tend to accumulate digital files, even when they are not directly relevant to my job	.482	-	.671	r=.835, p<.001
3. Deleting certain files would be like deleting a loved one	.902	r=.883, p<.001	<.3	-
4. If I delete certain files I feel apprehensive about it afterwards	.838	r=.918, p<.001	.373	-
5. I strongly resist having to delete certain files	.713	r=.908, p<.001	.492	-
6. I feel strongly that some files might be useful one day	<.3	-	.885	r=.813, p<.001
7. I lose track of how many digital files I possess	<.3	-	.857	r=.840, p<.001
8. Deleting certain files would be like losing part of myself	.876	r=.918, p<.001	<.3	
9. Thinking about deleting certain files causes me some emotional discomfort	.884	r=.866, p<.001	<.3	
10. At times I find it difficult to find certain files because I have so many	.556	-	.644	r=.841, p<.001

Table 1: The factor loadings for each question for factor 1 (difficulty deleting) and factor 2 (accumulating), small (<.3) factor loadings are supressed.

#### **TEST-RETEST**

Scale-to-scale correlations were conducted for the total score and there was a significant correlation from time 1 to time 2 (six weeks later) for difficulty deleting: r=.736, p<.001 and accumulating: r=.663, p<.001.

## DIGITAL BEHAVIOURS AT WORK QUESTIONNAIRE (DBWQ)

Data is reported for each of the four sections of the DBWQ, comparing those individuals reporting data protection responsibilities (DPR) with individuals reporting no data protection responsibilities (noDPR). Section 1 provided a quantitative assessment of the number of digital files currently stored and these are presented in Table 2.

A one-way ANOVA was conducted to examine the differences in the amount of files kept between those who perceived they had data protection responsibilities and those who did not. The DPR group displayed

significantly higher amounts of read emails in their inbox, unread emails in their inbox, presentation files, photographs, and total number of files compared to the noDPR group. Partial correlations controlling for group were then conducted to assess possible relationships between the number of files retained, and the two hoarding factors of accumulating and difficulty deleting. In every instance, a significant positive correlation was found between the number of each item retained and both hoarding factors (statistics available on request).

#### **TEST-RETEST**

There was a significant correlation from week one to six weeks later during the test- retest for this section in the sample as a whole (r=.644, p<.001).

Section 2 asked about deletion behaviours, specifically how often an individual deleted each type of file and deletion activity is represented in Table 3. A one-way ANOVA comparing the DPR and noDPR group revealed no significant differences in deletion activity for each type of file (full statistics available on request

	DPR Means (SD)	NoDPR Means (SD)	ANOVA
Read emails in inbox	359.3 (332.8)	254.0 (307.3)	F(1, 195)=4.030, p=.046*
Unread emails in inbox	128.5 (199.0)	29.8 (33.9)	F(1, 195)=13.095, p<.001**
Emails in 'deleted' folder	264.1 (294.5)	247.0 (343.6)	F(1, 195)=.121, p=.728
Emails in archived folders	304.8 (336.3)	269.6 (366.8)	F(1, 195)=.410, p=.523
Text files	257.6 (280.7)	224.9 (324.1)	F(1, 195)=.487, p=.486
Numerical files	212.7 (262.9)	156.1 (264.2)	F(1, 195)=1.785, p=.183
Presentation files	163.8 (230.2)	74.5 (195.5)	F(1, 195)=6.190, p=.014*
Photographs	215.7 (282.3)	116.7 (232.3)	F(1, 195)=5.211, p=.024*
Total number of files	1857.0 (1660.5)	1331.3 (1437.4)	F(1, 195)=.487, p=.039*

Table 2: Type and number of digital files associated with the workplace (DBWQ, Part 3, Section 1). Higher scores indicate higher amounts of files stored. This was scored on a continuous scale from 0–1000+, for those who perceived they had data protection responsibilities (DPR) compared to those who do not (noDPR).

<sup>\*</sup>Significant to .005, \*\*significant to <.001

from the report authors), and not surprisingly there were no significant differences between the groups in relation to total deletion scores in this section (F,202 = 1.56, p = 0.21). Subsequent bivariate correlations between deletion activity and the two hoarding factors thus comprise data from the whole sample and are displayed in Table 3.

#### **TEST-RETEST**

There was a significant correlation from week one to six weeks later during the test-retest for this section r=.759, p<.001.

Section 3 of the DBWQ picked up on the issue of email hoarding (as these were the most commonly hoarded files), asking for the reasons why participants kept their emails (Table 4). Here, a one-way ANOVA revealed

	Correlations with 'difficulty deleting'	Correlations with 'accumulating'
Read emails in inbox	r=.095, p=.180	r=.331, p<.001
Unread emails in inbox	r=-002 p=974	r=.127. p=071
Emails in 'deleted' folder	r=.171, p=.015	r=.357, p<.001
Emails in archived folders	r=136, p=.053	r=.183, p<.009
Text files	r=144, p=.041	r=.087, p=219
Numerical files	r=171, p=.015	r=.098, p=167
Presentation file	r=-207, p=.003	r=.062, p=.384
Photographs	r=224, p=.001	r=041, p=.560

Table 3: Participants were asked how often they delete digital files 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 files. These scores were correlated with the two DHQ factors and presented below with significant correlations in bold.

STATEMENT	DPR Total	NoDPR total	ANOVA
I don't delete them because they may come in useful in the future	5.0 (1.2)	4.6 (2.2)	F(1, 202)=2.034, p=.155
I don't delete them because they may contain information vital for my job	4.9 (1.7)	4.4 (3.3)	F(1, 202)=4.091, p=.044
I don't delete them in case I need to have 'evidence' that something has been done	4.4 (1.9)	3.4 (2.2)	F(1, 202)=7.216, p=.008
I don't delete them because I am worried that I might accidentally delete something important	4.1 (1.9)	3.3 (2.0)	F(1, 202)=7.236, p=.044
I don't delete them because I feel a sense of professional responsibility about them	3.8 (1.9)	2.8 (1. 9)	F(1, 202)=10.599, p=.001

Table 4: Top five reasons why participants keep emails. The questions are ranked in order of importance, using a seven-point Likert-type scale (1 = not at all, 7 = very much so) with mean scores and SDs (in parentheses) displayed, and one-way ANOVA comparing DPR and noDPR groups.

significant differences between the two groups in their degree of endorsement for most of the reasons for non-deletion. Table 4 thus shows the five most-commonly selected reasons for non-deletion as a function of group, and interestingly both groups endorsed exactly the same reasons in the same order of importance.

A partial correlation controlling for group was then conducted to assess the relationships between the ratings for each non-deletion statement and the two hoarding factors. All such correlations were significant and positive, with the exception of one (difficulty deleting was not significantly correlated with the statement 'I don't delete them because they may come in useful in the future, r = .059, p = 0.406).

The final section of the DBWQ asked for the perceptions of consequences if emails were released for the individual and their organisation, with the data being presented in Table 5, separated by group. A one-way ANOVA revealed a significant group difference for each statement, with the DPR group scoring higher than the noDPR group on each occasion. These data are interesting as it suggests that people are aware that the consequences of hoarding can be more severe for the organisation than for the individual, with significant differences existing between the implications for 'self' and 'company' for every file type (shown in Table 5). p<.001 respectively if digital files were released. A partial correlation controlling for group showed that both hoarding factors were significantly and positively

associated with each statement (full statistics available on request).

#### **TEST-RETEST**

There was a significant correlation from time 1 to time 2 for the total scores for 'consequences to themselves': r=.558, p<.001 and 'consequences to their company' r=.707, p<.001, additionally all other correlations were positive and significant (available on request).

File type	DPR	noDPR	ANOVA
Emails	3.8 (1.9)	2.6 (1.6)	F(1, 200) = 15.316, p<.001
Text files	3.8 (1.9)	2.51 (1.6)	F(1, 200) = 21.024, p<.001
Numerical files	3.9 (1.8)	2.53 (1.6)	F(1, 200) = 24.738, p<.001
Presentation files	3.5 (2.00)	2.24 (1.5)	F(1, 200) = 19.029, p<.001
Photographs	3.6 (2.00)	2.24 (1.5)	F(1, 200) = 19.414, p<.001

Table 5: Perceived consequences for participants if files were released (e.g. emailed to the wrong distribution group or stolen by a hacker and circulated on the internet) scored from no consequences at all (1) to very severe consequences (7) with mean scores and SDs (in parentheses) displayed. Table 5: Perceived consequences for participants if files were released (e.g. emailed to the wrong distribution group or stolen by a hacker and circulated on the internet) scored from no consequences at all (1) to very severe consequences (7) with mean scores and SDs (in parentheses) displayed.

#### **SAMPLE**

We used a purposeful sampling method to recruit 20 participants (15 females, 5 males). Participants were recruited from two large organisations in the North East of England; an academic organisation and a commercial organisation. Participants ranged from 18 to 59 years of age (Mean: 34.6 years SD: 10.1 years).

All participants were in either full-time or part-time employment, used a computer as part of their job, and had been with their company for at least six months. Participants were all identified as digital hoarders as they all scored highly on the Digital Hoarding Questionnaire (DHQ) (i.e. they had scores over 16 for difficulty discarding and 15 for accumulating). See Table 6 for each participant's professional job title and digital hoarding scores. Participants were each compensated a £10 Amazon gift voucher for taking part.

Participant – job title	Organisation	DBQ Score (Difficulty discarding/ accumulating)
P1: Senior Researcher	Academic organisation	DD:25 A:20
P2: Researcher	Academic organisation	DD:20 A:18
P3: PhD Researcher	Academic organisation	DD:17 A:16
P4: PhD Researcher	Academic organisation	DD:25 A:24
P5: Faculty Associate	Academic organisation	DD:20 A:18

P6: Senior Lecturer	Academic organisation	DD:16 A: 20
P7: Senior Lecturer	Academic organisation	DD:17 A:20
P8: Senior Lecturer	Academic organisation	DD:23 A:22
P9: Senior Researcher	Academic organisation	DD:20 A:16
P10: PhD Researcher	Academic organisation	DD:28 A:19
P11: PhD Researcher	Academic organisation	DD:16 A:20
P12: Associate Professor	Academic organisation	DD:22 A:36
P13: Senior Lecturer	Academic organisation	DD:28 A:32
P14: Demonstrator	Academic organisation	DD:24 A:28
P15: Senior Lecturer	Academic organisation	DD:19 A: 16
P16: PhD Researcher	Academic organisation	DD:20 A:18
P17: Research Chemist	Commercial organisation	DD:20 A: 18
P18: Senior Scientist	Commercial organisation	DD:18 A:24
P19: Intern Scientist	Commercial organisation	DD:20 A:18
P20: Research Scientist	Commercial organisation	DD:22 A:23

Table 6: Overview of participant's demographic information and digital hoarding scores.

#### MATERIALS AND PROCEDURE

The study received ethical approval prior to the study taking place. The study itself comprised two parts. In the first part participants completed a digital hoarding questionnaire, in the second part participants took part in focus groups. The results presented in this paper relate to the focus group data, but in order to understand how the participants for the focus groups were identified, information on the first part of the study is presented below.

The DHQ was sent around two organisations and individuals who scored highly were contacted and asked to take part in a focus group. The focus group schedule was formulated by creating open-ended and semi-structured questions grounded in current literature in order to keep on topic but allowing participants to provide further explanations and discuss their own experiences. Example questions on the interview schedule included: 'Thinking about those emails, 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; there were 2–5 participants in each focus group, however, one participant (P17) was interviewed individually. 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 took part on a voluntary basis. 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

We 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 reread several times. We noted the ways in which the transcripts contained evidence

of motivational and explanatory accounts of digital hoarding practices by our participants. We generated sets of initial codes that represented the different ways in which people accounted for, or simply described, their hoarding. We then started to group these together, identifying codes that were similar or explained the same aspect within the data were incorporated into a subtheme.

These subthemes were categorised into four underlying dimensions that made coherent sense and we felt captured the different motivations for hoarding behaviour that we saw in the data. We then reviewed the data to ensure that these categories were a good fit with the data, including checking for negative cases.

Finally, we defined and named each of the four dimensions and sought out example quotes from the transcripts to illustrate each dimension type. When we were happy with our four dimensions, we then examined the data with respect to a series of 'barriers to change'. These concepts were primarily driven by the small but existing literature on digital hoarding (Sweeten et al., 2018) but we were also open to identify new codes.

Our focus here was on providing a clear account of the ways in which digital hoarding behaviours are driven, justified and understood by our participants, and our approach relied on multiple collaborative coders to add meaning and complexity to the underlying typology. To those ends we did not use inter-coder reliability, as this implies a single or 'real' way of interpreting data, nor do we provide counts of the data.

#### **RESULTS**

Looking at the data we determined that there were underlying dimensions of digital hoarding in the workplace. 'Disengagement', 'Compliance', 'Anxiety' and 'Collection'. Then, we present our cross-cutting theme of barriers to change. The themes are illustrated with direct quotes from the transcripts.

Four underlying dimensions of digital hoarding within the workplace were identified in the transcripts. For most digital hoarders their workplace behaviour was driven by a primary dimension but for some several underlying dimensions in their behaviour were identified.

#### 1. DISENGAGEMENT



Disengagement characterised many of our participants' hoarding behaviours. Our hoarders had a lot of digital data stored in their email inboxes and in their folders. However, this data is not well organised.

Being disengaged with data management behaviours meant that participants expressed a lack of control over the digital data they acquire, as files are often sent to them unrequested and over time, these files have slowly accumulated almost without the hoarder realising.

'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 3,052 just in the read and what are they doing. You know, so, it is a complete shambles.' (P5, Faculty Associate).

This digital data has accumulated over time, and where this has been driven by a general sense of disengagement with the process of managing and structuring this data, many participants felt that it is now too late to try to organise it, and that they should have started good habits earlier on.

In addition, some hoarders are just too lazy or do not have the time to delete the information, and again this reflects the passive disengagement that drives their hoarding behaviour.

For these hoarders the data is not seen as being particularly valuable but nor is the hoarding behaviour seen as being problematic. These hoarding behaviours are not related. Participants were not concerned about the amount of digital data that they have only that it would just take too long to actually 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 opposed to close that notification and continue doing what I am doing, I don't see much point in it.' (P14, Demonstrator)

Hoarding behaviour that was underpinned by disengagement also led to individuals storing data on their personal storage systems. Participants described how it was easier to use external storage systems rather than organise or delete data. They also mentioned that they find it easier to access their data when it is stored on a cloud device.

Neave, Briggs, Sillence, McKellar

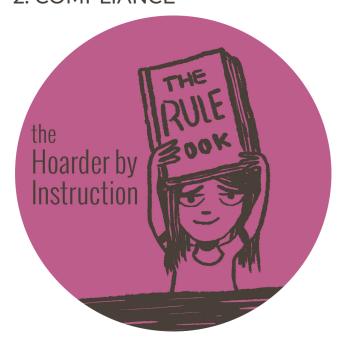
'I put the most important folders in my Dropbox so I can access them at work and at home.' (P2, Researcher)

These hoarders have lower ownership over the data they have stored and feel that much of it has been accumulated over time and almost accidentally. Deleting this data would not be problematic in the sense that a lot of it, especially emails, did not feel 'like theirs' but they are unlikely to delete data due to 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 Researcher)

The concept that the 'moment has passed' for organising 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 are ideas that we see echoed in the small but nascent body of work on digital hoarding (Thomas & Briggs, 2014; Vertesi et al., 2016).

#### 2. COMPLIANCE



For some hoarders, storing digital files and emails on work computers is driven by compliance. These participants described having a good knowledge and organisation of the emails and files they currently have stored and do so in order to comply with their company's instructions, guidance and policies. They keep information that they have been instructed to keep, or digital data that has been sent to them by their managers or colleagues.

Where the hoarding behaviour is underpinned by compliance, digital hoarders perceive little personal value in the digital data that 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 hoarders are happy with the amount of digital data they have stored and believe that it is 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)

If the data is no longer required, then people driven by compliance are also happy to delete information once they know that they no longer need to keep it. As these hoarders believe they need to keep all this digital data to be complaint they do not feel anxious or worried about the amount of information that they have. Therefore, they feel no anxiety over discarding information, as long as they have been told that they no longer have to keep it.

'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)

When hoarding behaviour is based on compliance the hoarders have lower ownership of the data they have stored and feel happy to share their information with

Cybersecurity risks of digital hoarding behaviours

other people. Data is kept on their work computer rather than external or cloud storage, unless they have a shared cloud system at their company. These hoarders keep a lot of their information on shared drives, as they are happy to share data with other people, as they do not see it as their own personal data but data they keep 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 organisations, 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 hoarders in our study, keeping data was seen as part of doing their job and being efficient in how they managed their work. Whilst hoarding is typically associated with individual factors, hoarding that is driven by compliance may well still lead to sharing behaviours especially in situations involving team or collaborative working environments.

3. ANXIETY



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 organise 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, PhD Researcher)

Participants recognised that many of the emails they kept were no longer valuable but were still reluctant to delete them in case they might need them as evidence in the future.

'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)

This finding resonates with Sweeten et al. (2018) who found that keeping digital data as 'evidence' was important in personal settings, for instance electronic receipts or delivery notices. Keeping data even if its value is unclear to participants provides a level of security and comfort for digital hoarders.

Participants were also anxious that information could be become important for their job or for the organisation at some future point in time and therefore the best plan was to keep everything just in case. The only deletions possible were in the cases were obvious spam emails, clearly not associated with them personally or 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

leave, Briggs, Sillence, McKellar

comes up and I have to remind myself about something.' (P10, PhD Researcher)

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

'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)

4. COLLECTION



For some participants hoarding large amounts of digital data was something that was a purposeful decision. Where the underlying dimension of hoarding was collection 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; they were typically systematically organised into folders and were rarely deleted. The only emails considered for deletion were obvious spam emails (for example, phishing emails).

Again, hoarders felt that most emails may be valuable and needed again in the future.

'Mine is the same, I don't delete work emails, unless it is like the fake journal emails that come through.' (P11, PhD Researcher)

'I actually think you never know how valuable something is going to be, until it might at some point become valuable.' (P12, Associate Professor)

Collection drove hoarding behaviours that led to the collection of purposely stored, well-organised digital files. These hoarders have well-labelled files and know exactly what files are stored and where the file is within their filing system. If their workplace has storage limits they get 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 U drives 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 is an important motivation, there are few if any concerns about the amount of digital data that is being stored. This is data is seen as valuable to hoarders now or will 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. Hoarding a valuable collection of digital data meant that these hoarders saw the stored digital data as part of their identity and felt a sense of pride from having all this digital data.

These hoarders were happy that other colleagues knew that they kept a lot of data and would often call on them to send information.

Cybersecurity risks of digital hoarding behaviours

'I feel fine about the amount of data I have, you never know when you might need it.' (P20, Research Scientist)

'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 hoarders 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. As with previous work on digital artefacts, participants displayed a strong sense of ownership over their digital possessions (Cushing, 2013).

'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)

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

#### BARRIERS TO CHANGE

Most participants believed that there were some advantages to keeping large amounts of digital data. This is because within their workplaces many had unlimited storage space, therefore, there were no reason to delete any data in order to free up space. If they did reach any storage limits, participants are happy 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 is different from keeping physical data, for example paper documents that can 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 is relatively easy to find digital files, even if they have 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 receive very little guidance on data deletion and retention. Most of the participants said they did not know their organisations retention and deletion policies, despite having recent GDPR training. Some participants mentioned that their company hosts data management days, however, as this is left up to each individual to do, they are still not sure which files should be retained or deleted. Most participants mentioned 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 Researcher)

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 organisation rather for them as individuals. Where they did imagine some kind of data loss, some were dismissive, arguing that if everyone was 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

Neave, Briggs, Sillence, McKellar

devices. For example, participants discussed the implications of having their laptop or memory stick stolen, but ironically, perceived solutions to this problem was the duplication of data, i.e. ensuring backups existed on other systems.

Overall, participants saw the benefits to keeping large amounts of digital data, recognising that 'unlimited' storage coupled with good search facilities meant there were no downsides. If they did foresee any issue with security breaches, these were seen as generating consequences for the company rather than for themselves.

### **REFERENCES**

Adom, W., Sellen, A., Harper, R., & Thereska. E. (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.

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

Bergman, O., and Beyth-Marom, R. (2003). The user-subjective approach to personal information management systems, *Journal of the American Society for Information Science and Technology*, 54: 872-878.

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).

Boardman, R., and Sasse, M.A. (2004). Stuff goes into the computer and doesn't come out. A cross-tool study of personal information management. *Proceedings of the SIGCHI conference*, Vienna, 24<sup>th</sup>-29<sup>th</sup> April, vol 6 number 1.

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

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*: 523-548.

CGOC Information Governance Process Maturity Model, January, (2017). Accessed, 25th January, 2017 from: <a href="http://www.cgoc.com/">http://www.cgoc.com/</a>

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

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: 1723-1734.

Dabbish, L. A., Kraut, R. E., Fussell, S., and Kiesler, S. (2005). Understanding email use: predicting action on a message. In *Proceedings of the SIGCHI conference on Human factors in computing systems:* 691-700.

Evans, J.M., Hendron, M.G., and Oldroyd, J.B. (2014). Withholding the ace: the Individual- and unit-level performance effects of self-reported and perceived knowledge hoarding. *Organisation Science* 26.2: 494-510.

Frost, R., and Gross, R. (1993). The hoarding of possessions. *Behaviour Research and Therapy, 31*: 367-381.

Gerritsen, D.B., Tasse, D., Olsen, J.K., Vlahovic, T.A., Gulotta, R., Odom, W., & Zimmerman, J. (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., and Gormley, S.J. (2012). Data hoarding and information clutter: the impact of cost, life span of data, effectiveness, sharing, productivity and knowledge management culture. *Issues in Information Systems*, 13: 90-95.

Gulotta, R., Odom, W., Forlizzi, J., and Faste, H. (2013). Digital artifacts as legacy: exploring the lifespan and value of digital data. *CHI 2013: Changing Perspectives, Paris*: 1813-1822.

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

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

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.

Lansdale, M. (1988). The psychology of personal information management. *Applied Ergonomics*, 19: 55-66.

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.

Oravec, J. A. (2015). Depraved, distracted, disabled, or just pack rats? Workplace hoarding persona in physical and virtual realms. *Persona Studies*, *1*, 75-87.

Oravec, J.A. (2017). Digital (or virtual) hoarding: emerging implications of digital hoarding for computing, psychology, and organisation science. *International Journal of Computers in Clinical Practice*, 3: 27-39.

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

Sprague, R. (2008). Orwell was an optimist: the evolution of privacy in the United States and its de-evolution for American employees. *John Marshall Law Review*, 42, 83-134.

Steketee, G., and Frost, R. (2003). Compulsive hoarding: current status of the research. *Clinical Psychology Review*, 23: 905-927.

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

van Bennekom, M.J., Blom, R.M., Vulink, N., and Denys, D. (2015). A case of digital hoarding. *British Medical Journal Case Reports*, doi: 10.1136/bcr-2015-210814.

Vertesi, J., Kaye, J., Jarosewski, S.N., Khovanskaya, V.D., & Song, J. (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.

Vitale, F., Janzen, J., and McGrenere, J. (2018). Hoarding and minimalism: tendencies in\_digital\_data preservation. *CHI 2018*, *April 21-26*, *Montreal QC*, *Canada*.

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

PART 1: AROUT YOU

## APPENDIX: DIGITAL BEHAVIOURS QUESTIONNAIRE (DBQ)

We now use and experience many different types of digital information with different formats and on different devices – e.g., apps, data files, web links, blogs, photos etc. We are interested in how you behave towards, and how you think about the different types of digital information you have access to in your daily life and in your workplace. Please answer the following questions as honestly as you can, if you prefer not to give an answer then leave it blank.

. ,	,			
1.1	I am: mal	е 🗆	female □	prefer not to say $\square$
1.2	I am years o	ld		
1.3	My current emplo	yment s	tatus is:	
	Working part-time Working full-time			
1.4	Current employer	:		
1.5	Size of current en	ployer:		
	Micro enterprise (< Small enterprise (< Medium enterprise Large organisation	50 staff) (<250 s	taff)	
1.6	Length of time wi	th curre	ent employer:	
	Less than 1 year 1-5 years 5-10 years Over 10 years			
1.7	Current job role			
1.8	Length of time in	your cu	rrent job role	:
	Less than 1 year 1-5 years 5-10 years Over 10 years			

### APPENDIX: DIGITAL BEHAVIOURS QUESTIONNAIRE (DBQ) Neave, Briggs, Sillence, McKellar

1.9	Does	your o	current j	job role entai	l any	responsibility	for data	protection?
	Yes		No					

#### PART 2: HOW YOU FEEL ABOUT DIGITAL INFORMATION IN GENERAL

We are interested in how people feel about digital materials in their workplace. These materials which we refer to as 'files', include emails, email attachments, spreadsheets, PDF's, databases 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

	Noi	at all		Very much so				
2.1. I find it extremely difficult to delete old or unused files	1	2	3	4	5	6	7	
2.2. I tend to accumulate digital files, even when they are not directly relevant to my job	1	2	3	4	5	6	7	
2.3. Deleting certain files would be like deleting a loved one	1	2	3	4	5	6	7	
2.4. If I delete certain files I feel apprehensive about it afterwards	1	2	3	4	5	6	7	
2.5. I strongly resist having to delete certain files	1	2	3	4	5	6	7	
2.6. I feel strongly that some files might be useful one day	1	2	3	4	5	6	7	
2.7. I lose track of how many digital files I possess	1	2	3	4	5	6	7	
2.8. Deleting certain files would be like losing part of myself	1	2	3	4	5	6	7	
2.9. Thinking about deleting certain files causes me some emotional discomfort	1	2	3	4	5	6	7	
2.10. At times I find it difficult to find certain files because I have so many	1	2	3	4	5	6	7	

#### PART 3: ABOUT YOUR DIGITAL BEHAVIOUR AT WORK

We are now interested in the range of digital materials that people typically have access to in their working 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 emails, email attachments, spreadsheets, PDF's, databases etc. When you answer, do not consider spam/junk files.

#### APPENDIX: DIGITAL BEHAVIOURS QUESTIONNAIRE (DBQ)

Cybersecurity risks of digital hoarding behaviours

#### SECTION 1: ACCUMULATION AND STORAGE BEHAVIOURS

Below is a list of common digital items you might currently have stored on your work computer/network drive. For each one, please indicate how many you have right now. If you have access to your electronic devices 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 file:
Read emails currently in inbox
Unread emails currently in inbox
Emails currently in 'deleted' folder
Emails in archived folders:
Text files:
For example, word documents, reports, PDF's etc.
Numerical files:
For example, statistical data files, spreadsheets, databases etc.
Presentation files:
For example, PowerPoint files, poster files etc.
Photographs:
CECTION 2. DELETION DELIAMONDO

#### **SECTION 2: DELETION BEHAVIOURS**

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.

## APPENDIX: DIGITAL BEHAVIOURS QUESTIONNAIRE (DBQ) Neave, Briggs, Sillence, McKellar

File Type	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
Read emails currently in inbox					
Unread emails currently in inbox					
Emails currently in 'deleted' folder					
Emails in archived folders					
Text files. For example, word documents, reports, PDF's etc.					
Numerical files. For example, statistical data files, spreadsheets databases etc.					
Presentation files. For example, PowerPoint files, poster files etc.					
Photographs					

#### **SECTION 3: RATIONALE FOR KEEPING E-MAILS**

Now think specifically about the emails you keep (in your inbox, or in archived folders). If you rarely delete them, can you identify the key reasons why not? For each of the following statements please indicate how typically true this is for you, where 1 = not at all true, and 7 = very true.

	Not at all true				Very true			
a) It is my company policy never to delete information so I don't have a choice	1	2	3	4	5	6	7	
b) I don't delete them because they may come in useful in the future	1	2	3	4	5	6	7	
c) I don't delete them because they may contain information vital for my job	1	2	3	4	5	6	7	
d) I don't delete them because I am worried that I might accidentally delete something important	1	2	3	4	5	6	7	
e) I don't delete them because I feel a sense of attachment to them	1	2	3	4	5	6	7	

#### APPENDIX: DIGITAL BEHAVIOURS QUESTIONNAIRE (DBQ)

Cybersecurity risks of digital hoarding behaviours

f) I don't delete them because I feel a sense of professional responsibility about them	1	2	3	4	5	6	7
g) I don't delete them because they 'belong' to my company and are not mine to do with as I wish	1	2	3	4	5	6	7
h) I don't delete them because storing them is not my problem, if they take up too much space then my company can delete them	1	2	3	4	5	6	7
i) I simply don't have the time to delete them all	1	2	3	4	5	6	7
j) I am too lazy to delete them	1	2	3	4	5	6	7
k) I don't delete them in case I need to have 'evidence' that something has been done	1	2	3	4	5	6	7
l) I don't delete them because I keep an example from everyone so that it is easier to reply in future	1	2	3	4	5	6	7

#### **SECTION 4: CONSEQUENCES**

For each of the types of files you may have stored on your work computer/network drive/external drives please now consider the degree of sensitivity of that material and the possible consequences if it were made public or stolen. Think about if the files were to be released (e.g., emailed to the wrong distribution group, or stolen by a hacker and circulated on the internet). Think firstly about consequences for you, and then secondly about consequences for your company.

In the scale selecting 1 = no consequences at all, whilst 7 = very severe consequences.

#### Personal consequences for me...

	No consequences					Very severe			
Emails	At a	2	3	4	5		sequen 7	ces	
Text files	1	2	3	4	5	6	7		
Numerical files	1	2	3	4	5	6	7		
Presentation files	1	2	3	4	5	6	7		
Photographs	1	2	3	4	5	6	7		

## APPENDIX: DIGITAL BEHAVIOURS QUESTIONNAIRE (DBQ) Neave, Briggs, Sillence, McKellar

#### Consequences for my company...

	No consequences					Very severe		
Emails	At a		3	4	5		rsequences 7	I
Text files	1	2	3	4	5	6	7	
Numerical files	1	2	3	4	5	6	7	
Presentation files	1	2	3	4	5	6	7	
Photographs	1	2	3	4	5	6	7	

For more information on CREST and other CREST resources, visit www.crestresearch.ac.uk

