

Digital hoarding behaviours: measurement and evaluation

The social and psychological characteristics of individuals who hoard physical items are quite well understood, however very little is known about the psychological characteristics of those who hoard digital items and the kinds of material they hoard. In this study, we designed a new questionnaire (Digital Behaviours Questionnaire: DBQ) comprising 2 sections: the Digital Hoarding Questionnaire (DHQ) assessing two key components of physical hoarding (accumulation and difficulty discarding); and the second measuring the extent of digital hoarding in the workplace (Digital Behaviours in the Workplace Questionnaire: DBWQ). In an initial study comprising 424 adults we established the psychometric properties of the questionnaires. In a second study, we presented revised versions of the questionnaires to a new sample of 203 adults, and confirmed their validity and reliability. Both samples revealed that digital hoarding was common (with emails being the most commonly hoarded items) and that hoarding behaviours at work could be predicted by the 10 item DHQ. Digital hoarding was significantly higher in employees who identified as having 'data protection responsibilities', suggesting that the problem may be influenced by working practices. In sum, we have validated a new psychometric measure to assess digital hoarding, documented some of its psychological characteristics, and shown that it can predict digital hoarding in the workplace.

Keywords: Digital hoarding; saving behaviours, difficulty discarding; questionnaire development; personal information management.

INTRODUCTION

The demographic, social, and personality characteristics of individuals who hoard physical objects have been extensively investigated from psychological, clinical and sociological perspectives (see Frost and Gross 1993; Grisham and Barlow 2005; Steketee et al. 2003; Nordsletten et al. 2013; Tolin 2011). Recently, there has been much speculation in the media 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). While 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). While 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, organize 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 organizations, 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. These authors noted that the ability to properly identify and quantify digital hoarding behaviours was currently lacking, and so the aim of this current study was to develop and validate a psychometric questionnaire to identify digital hoarders (the Digital Hoarding Questionnaire - DHQ) and an associated questionnaire to measure patterns of digital hoarding behaviours within the workplace (Digital Behaviours in the Workplace Questionnaire: DBWQ). The latter would ensure that we gained an understanding of the scale of the problem and the potential consequences for both the organisation and individual. We would expect the DHQ to be predictive of such workplace behaviours.

Study 1

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). In terms of data protection responsibilities, 418 respond to this question with 278/418 (66.5%) admitting such responsibilities. Participants were obtained through a market research company, 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 DHQ comprised a series of 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 (henceforth the Digital Behaviours in the Workplace Questionnaire or DBWQ) was created as a way 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 our 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.

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 a market research company (Critical Mix – www.criticalmix.co.uk). 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

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 6 items (Cronbach's alpha = .905) which we call 'difficulty deleting' and the other comprising 4 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 (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 (items 4 and 9 shown in table 1) were found to be associated with the factor of 'difficulty deleting'. These two factors were strongly positively correlated with one another ($r = .740, p < .001$), and to hoarding behaviours at work as measured by the DBWQ, specifically difficulty deleting with the total number of digital files retained ($r = .416, p < .001$), and accumulating with the total number of digital files ($r = .416, p < .001$).

This pilot 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 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.

Study 2

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 the 203 individuals, 148 participants (72.9%) stated that they had data protection responsibilities.

Materials

The final versions of both questionnaires are provided in the appendix and comprise, firstly, the revised ten-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 5 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 twelve 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 7-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 one and two, 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 7-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 (see below).

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. A random sample of 50 individuals were asked to re-take the study again 6 weeks after first taking part so that we could establish the test-retest reliability of the scale. Scale to scale correlations were conducted for the total score and there was a significant correlation from time 1 to time 2 (6 weeks later) for difficulty deleting: $r=.736$, $p<.001$ and accumulating: $r=.663$, $p<.001$.

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 6 items (Cronbach's alpha = .945) which we call 'difficulty deleting' and the other comprising 4 items (Cronbach's alpha = .873) which we call 'accumulating'. Both scales correlated substantially ($r = .756$).

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

| | 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 |

Digital Behaviours at Work Questionnaire (DBWQ)

In the test-retest, we found a significant correlation from week 1 to 6 weeks later during the test re-test for this section in the sample as a whole (r=.644, p<.001). 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.

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 compared to those who do not.

| | DDPR Means (SD) | NoDP R Means (SD) | ANOVA | Partial correlations for difficulty deleting | Partial correlations for accumulating |
|----------------------------|----------------------------|--------------------------------------|--------------------------------------|---|--|
| Read emails in inbox | 359.3 (332.8) | 254.0 (307.3) | F(1, 195)=4.03 0, p=.046* | r=.149, p.037* | r=.290, p<.001** |
| Unread emails in inbox | 128.5 (199.0) | 29.8 (33.9) | F(1, 195)=13.0 95, p<.001** | r=.334, p<.001** | r=.334, p<.001** |
| Emails in 'deleted' folder | 264.1 (294.5) | 247.0 (343.6) | F(1, 195)=.121, p=.728 | r=.289, p<.001** | r=.332, p<.001** |
| Emails in archived folders | 304.8 (336.3) | 269.6 (366.8) | F(1, 195)=.410, p=.523 | r=.195, p=.008* | r=.336, p<.001** |
| Text files | 257.6 (280.7) | 224.9 (324.1) | F(1, 195)=.487, p=.486 | r=.264, p<.001** | r=.381, p<.001** |
| Numerical files | 212.7 (262.9) | 156.1 (264.2) | F(1, 195)=1.78 5, p=.183 | r=.298, p<.001** | r=.351, p<.001** |
| Presentation files | 163.8 (230.2) | 74.5 (195.5) | F(1, 195)=6.19 0, p=.014* | r=.339, p<.001** | r=.316, p<.001** |
| Photographs | 215.7 (282.3) | 116.7 (232.3) | F(1, 195)=5.21 1, p=.024* | r=.282, p<.001** | r=.349, p<.001** |
| Total number of files | 1857.0 (1660.5) | 1331.3 (1437.4) | F(1, 195)=.487, p=.039* | r=.354, p<.001** | r=.459, p<.001** |

*Significant to .005, **significant to <.001

A one-way ANOVA was conducted to examine the differences in amount of files kept between those who perceived they had data protection responsibilities (DPR) and those who did not (noDPR). 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 in table 2).

Section 2 asked about deletion behaviours, specifically how often an individual deleted each type of file and deletion activity is represented in table 3. In the test-retest there was a significant correlation from week 1 to 6 weeks later ($r=.759$, $p<.001$). A one-way ANOVA comparing the DPR and no DPR group revealed no significant differences in deletion activity for each type of file (full statistics available on request), and not surprisingly there were no significant differences between the groups in relation to total deletion scores in this section ($F_{2,202} = 1.56$, $p = 0.21$). Subsequent bivariate correlations between deletion activity and the two hoarding factors thus comprise data from the whole sample and are displayed in table 3.

Table 3: In the DBWQ 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**.

| | 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 |

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 significant differences between the two groups in their degree of endorsement for most of the reasons for non-deletion. Table 4 thus shows the 5 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.

Table 4: Top 5 reasons why participants keep emails, the questions are ranked in order of importance, using a 7-point Likert-type scale (1 = not at all, 7 = very much so) with mean scores and SD's (in parentheses) displayed, and 1-way ANOVA comparing DPR and noDPR groups.

| STATEMENT | DP Total | Non-DP 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 |

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. In the 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).

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.

Table 5: Perceived consequences for participants if files were released (e.g. emailed to the

| | 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 |

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 SD's (in parentheses) displayed.

A partial correlation controlling for group showed that both hoarding factors were significantly and positively associated with each statement (full statistics available on request).

DISCUSSION

The aim of this project was to develop a psychometrically-valid questionnaire to measure digital hoarding behaviours in the workplace and to see if such behaviours are associated with the known characteristics of physical hoarding. In an initial study we developed the Digital Behaviours Questionnaire (DBQ) comprising two sections - The Digital Behaviours at Work Questionnaire (DBWQ) and the Digital Hoarding Questionnaire (DHQ). In a large sample of employees the DBWQ was found to provide an accurate assessment of digital hoarding behaviours, and showed good evidence of reliability (average test-retest of .677, internal consistency etc) and clearly distinguished between those with and without data protection responsibilities. However, we acknowledge that concurrent validity (e.g., comparisons with other measures of similar and divergent constructs) was not addressed and future studies need

to assess this new questionnaire in different samples, and compare with existing measures, such as questionnaires known to accurately assess physical hoarding behaviours.

The DHQ assesses two key factors associated with physical hoarding. The first factor, difficulty deleting, evokes feelings of loss or distress when data is deleted and seem to relate to the more emotional aspects of hoarding that are discussed in both the traditional hoarding literature (where ‘difficulty discarding’ is a known component of hoarding in both the Coles et al. 2003 and Tolin et al. 2010 studies); and in the relatively sparse digital hoarding work (e.g. Vitale et al. 2018 description of the emotional component to digital hoarding). In contrast, the second factor ‘accumulation’ suggests that the mass collection of digital files is simply perceived as the more practical and low-effort solution to the management of data at scale. Again, this would seem to relate to the ‘acquisition’ factor present in traditional hoarding studies (e.g. Coles et al. 2003; Tolin et al. 2010) and the ‘practical’ component of Vitale et al. (2018) study of virtual hoarding.

After appropriate revision the DBQ was then re-tested in a different sample of employees with the test-retest statistics showing an average correlation of .677 for the different sections. In this second study we asked employees whether or not they held data protection responsibilities, with around 70% stating that they did. The performance of these two groups were then compared on the questionnaire. Section 1 comprises information about the number of digital files and here the DPR group retained significantly more read emails, unread emails, presentation files and photographs than the noDPR group. This is not surprising as individuals with data protection responsibilities are perhaps more likely to receive extra files in response to workplace requirements. However the groups did not differ in number of emails in deleted or archived folders, and after controlling for group, partial correlations revealed significant positive correlations for each file type and the hoarding factors of ‘accumulation’ and

‘difficulty deleting’. This suggests that while job requirements may influence data gathering and storage, personality factors associated with hoarding behaviours are also important.

Section 2 asked about how often an individual deleted each type of file and here there was no significant group difference on any file type. Correlations were conducted between each file type and the two hoarding factors, and here the pattern is a little more nuanced with differences between the two DHQ factors emerging. Firstly, ‘accumulators’ seem reluctant to delete emails, while those having ‘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 participants 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 loathe to delete emails.

Section 3 focussed specifically on emails (as these were the most commonly retained files for both groups), asking for the reasons why emails were not deleted. Some group differences emerged in relation to the strength of endorsement for each reason for some of the reasons participants kept their emails with the DPR group expressing higher endorsement ratings, but interestingly both groups endorsed exactly the same reasons in the same order. Once more this appears to reflect the workplace requirements affecting more those individuals with data protection responsibilities. Partial correlations controlling for group demonstrated significant positive correlations between each reason and the two hoarding factors, 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”). This could reflect the fact that ‘accumulating’ represents the more functional goal of hoarding emails just in case they might come in useful, but show no correlation with the more emotional ‘difficulty deleting’, suggesting emails are seen as having little personal value. It is also worth pointing out that the third ranked reason for keeping emails, and one that correlates with both factors on the DHQ, is that there might be a need to supply ‘evidence’ at some future point, which relates to an interesting literature on the way individuals store digital files as a means to establish their own power base within an organisation (Gormley and Gormley, 2012).

The final section asked for the perceptions of consequences if files were released for the individual and their organisation. Once more, the DPR group scored significantly higher than the noDPR group on every statement, and again this likely reflects their greater awareness of the dangers relating to data protection. Again, after controlling for group the two hoarding factors were found to be significantly positively associated with each statement.

The DBWQ 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 organization 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. The implication is that they are hoarding data because of the requirements associated with their job, but there may be more to this relationship as while we might expect the DPR group to have more data, there is no reason why they should

then retain it. In fact, in light of their perhaps more specialist knowledge relating to data protection, we might indeed expect them to delete more data! The partial correlations controlling for group showed significant positive associations between the majority of responses and the two hoarding factors of accumulating and difficulty deleting, suggesting that data retention is also driven by the personality characteristics of the employee. It is possible for example 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. 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 now 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.

This work sits within a small but recent literature about digital hoarding behaviours and informs a larger literature on information management. To date, the digital hoarding studies that have been reported are qualitative investigations that have recognised the existence of extremes in the retention of digital files. For example, in the Vitale et al. (2018) study, both people who hoarded and minimalists were identified and interviewed. Those who showed hoarding behaviours described beliefs and behaviours that resonate well with our findings. For example, they discussed the more emotional elements of hoarding in terms of feeling distress at the thought of ‘letting go’ certain files and these tended to be files with more personal value such as photographs. On the other hand, they also discussed the tendency to keep files because the

might come in useful at some point in the future – behaviour more aligned with our ‘accumulating’ factor.

There is a strong sense from this small literature that people perceive these hoarding behaviours as harmless, fuelled in part by the fact that digital storage is cheap and search engines are fast, meaning that they perceive the costs of a ‘default’ decision to store files as low. Yet in the wider information management literature, the accumulation of potentially sensitive organisational data, stored in personal repositories, can be problematic. For example, both Evans et al. (2014) and Gormley & Gormley (2012) describe the ways that individuals accumulate organisational information in strategic ways, both retaining information that may be of future use to themselves (and thereby creating a false sense of ‘uncertainty avoidance’) but also withholding information from others in order to enhance their power-base.

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. the 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 involved increased surveillance resulting in workplace privacy concerns. 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.

This work forms a starting point towards the better understanding of workplace hoarding behaviours, and there are some limitations with this study. We used a market research company to recruit our participants and were thus unable to fully describe the sampling strategy and full demographic characteristics of our sample. In addition, the samples did not specify particular types of employees, job roles or types of organizations, and such variables may have a large impact upon the behaviours observed in the samples. Different organizations and job roles may impact upon the opportunity for digital hoarding, for example some organizations have mandatory email deletion policies, and so the questionnaire may not be applicable or useful in certain occupational settings. Digital hoarding behaviours could of course have an impact outside of the workplace, many people are 'freelance' or work at 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.

An additional limitation could be the reliance upon focussing on emails, which was justified as such files were the most commonly hoarded type of file, but other types of files could be more problematic in different organisations. A final limitation relates to the self-reported nature of the data relating to the number of files currently stored. Respondents were asked to use a slider ranging from 0-1000+, which is rather granular and may lead to inaccuracies in reporting. In future versions of the questionnaire (and in the version cited in the appendix) this has been changed to now ask for an exact number where possible, or to try and give an accurate estimate. It is of course now possible for individuals to gain a precise indication of the number and type

of digital files by using software such as ‘Cardinal’ (Dinnen et al. 2016) and incorporating such accuracy may enhance future work in this field.

REFERENCES

- 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.
- 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, 24th-29th April, vol 6 number 1*.
- CGOC Information Governance Process Maturity Model, January, (2017). Accessed, 25th January, 2017 from: <http://www.cgoc.com/>
- Coles, M.E., Frost, R.O., Heimberg, R.G., and Steketee, G. (2003). Hoarding behaviours in a large college sample. *Behaviour Research and Therapy*, 41: 179-194.
- 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.
- Dinnen, J.D., Odoni, F., Frissen, I., & Julien, C-A. (2016). Cardinal: novel software for studying file management behaviour. *Proceedings of the 79th ASIS&T Meeting: Creating knowledge, enhancing lives through information & technology*. American Society for Information Science, Silver Springs, MD, article 62, 10 pages.
- 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. *Organization Science* 26.2: 494-510.
- Frost, R., and Gross, R. (1993). The hoarding of possessions. *Behaviour Research and Therapy*, 31: 367-381.
- 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.
- Grisham, J.R., and Barlow, D.H. (2005). Compulsive hoarding: current research and theory. *Journal of Psychopathology and Behavioral Assessment*, 27: 45-52.
- 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.
- 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.
- 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.
- Steketee, G., Frost, R., and Kyrios, M. (2003). Cognitive aspects of compulsive hoarding. *Cognitive Therapy and Research*, 27: 463-479.
- Sweeten, G., Sillence, E., and Neave, N. (2018). Digital hoarding behaviours: underlying motivations and potential negative consequences. *Computers in Human Behavior*, 85: 54-60.
- Tolin, D.F. (2011). Understanding and treating hoarding disorder: a biopsychosocial perspective. *Journal of Clinical psychology: In Session*, 67: 517-526.
- Tolin, D.F., Frost, R.O., and Steketee, G. (2010). A brief interview scale for assessing compulsive hoarding: the Hoarding Rating Scale. *Psychiatry Research*, 30: 147-152.
- 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.
- Vitale, F., Janzen, J., and McGrenere, J. (2018). Hoarding and minimalism: tendencies in digital_ data preservation. *CHI 2018, April 21-26, Montreal QC, Canada*.

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.

PART 1: ABOUT YOU

1.1 I am: male female prefer not to say

1.2 I am ____ years old

1.3 My current employment status is:

Working part-time

Working full-time

1.4 Current employer: _____

1.5 Size of current employer:

Micro enterprise (<10 staff)

Small enterprise (<50 staff)

Medium enterprise (<250 staff)

Large organization (>250 staff)

1.6 Length of time with current 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 current job role:

Less than 1 year

1-5 years

5-10 years

Over 10 years

1.9 Does your current job role entail 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.

THE DIGITAL HOARDING QUESTIONNAIRE

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

| | <i>Not at all</i> | | | <i>Very much so</i> | | | |
|---|-------------------|---|---|---------------------|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 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.

THE DIGITAL BEHAVIOURS AT WORK QUESTIONNAIRE.

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:

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.

| 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. <i>For example, word documents, reports, PDF's etc.</i> | | | | | |
| Numerical files. <i>For example, statistical data files, spreadsheets databases etc.</i> | | | | | |
| Presentation files. <i>For example, PowerPoint files, poster files etc.</i> | | | | | |
| 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.

| | <i>Not at all true</i> | | | | <i>Very true</i> | | |
|---|------------------------|---|---|---|------------------|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 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 |
| 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, while 7 = very severe consequences.

Personal consequences for me...

| | <i>No consequences At all</i> | | | | | <i>Very severe consequences</i> | |
|--------------------|-----------------------------------|---|---|---|---|-------------------------------------|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Emails | | | | | | | |
| Text files | | | | | | | |
| Numerical files | | | | | | | |
| Presentation files | | | | | | | |
| Photographs | | | | | | | |

Consequences for my company...

| | <i>No consequences At all</i> | | | | | <i>Very severe consequences</i> | |
|--------------------|-----------------------------------|---|---|---|---|-------------------------------------|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Emails | | | | | | | |
| Text files | | | | | | | |
| Numerical files | | | | | | | |
| Presentation files | | | | | | | |
| Photographs | | | | | | | |