Digital accumulation behaviours and information management in the workplace: Exploring the tensions between digital data hoarding, organisational culture and policy

Kerry McKellar[1*], Elizabeth Sillence[1], & Nick Neave[2] & Pam Briggs[1]

1 Psychology & Communication Technologies, Department of Psychology, Newcastle upon Tyne, NE1 8ST, United Kingdom.

2 Hoarding Research Group, Department of Psychology, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom.

*corresponding author: Department of Psychology, Faculty of Health & Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom. Tel.: +44 (0)191 227 3716 Kerry Lakey E-mail address: Kerry.l.lakey@northumbria.ac.uk

Abstract

Individuals within organisations necessarily hold data, some of it containing personal identifiable data. For those individuals with a tendency to accumulate digital data and a reluctance to delete it, the potential for data to be stored (and thus be at risk) is greater. Understanding more about why people engage in digital data hoarding, whether they recognise the data they keep and how they respond to the mitigations put in place is important. Eleven people (Men= 9) working in a large commercial organisation who scored highly on a digital hoarding questionnaire, took part in focus groups to understand the extent to which they kept digital data, including personal digital data. The focus groups also explored employee compliance with policies and procedures including knowledge of GDPR. Thematic analysis led to three themes: (1) Organisational culture versus digital hoarding tendency, (2) Thinking about personal data and (3) Responsibility and control. The findings highlight different motivations for data hoarding including anxiety driven by 'blame culture' as participants respond to workplace challenges and the need to be accountable. Participants attended training and were aware of policies including GDPR but often used workarounds to keep data longer than specified in their information retention policies. Technical approaches to data reduction were also sometimes counterproductive. Findings are discussed in relation to the potential threat of digital data hoarding and technical and non-technical approaches to reducing digital data retention.

Keywords: digital hoarding, data accumulation, information retention policy, information management, workplace

Introduction

Digital data hoarding

Digital data hoarding is starting to be recognised as an important research topic (Oravec, 2017, 2018; Luxon, Hamilton, Bates & Chasson, 2019) with a greater focus on the ways in which people accumulate data at scale and/or struggle to delete digital data files. Digital hoarding has been defined as "…the acquisition of and failure to discard digital content, leading to the accumulation of digital clutter**"** (Sedera & Lokuge, 2018). A small but nascent literature on digital hoarding behaviours has examined motivations and potential negative consequences including reduced productivity and feelings of anxiety in relation to the accumulation of data (Sedera & Lokuge, 2018; Sweeten, Sillence & Neave, 2018; Vitale, Janzen & McGrenere, 2018; van Bennekom, Blom, Vulink & Denys, 2015). Studies examining digital

hoarding behaviours in more detail have tended to examine the data management practices of the general public across a range of settings rather than focusing specifically on the workplace. For example, Sweeten et al. (2018) asked individuals about their digital hoarding behaviours, motivations and the consequences of their data management practices. The study identified five main barriers to data deletion including keeping it for the future/or just in case, keeping it as evidence and emotional attachment.

Two recent studies have focused on digital hoarding behaviours in the workplace and are relevant here. In a workplace survey, Neave, McKellar, Briggs & Sillence (2019) found that digital accumulation was common and those individuals at the extreme of the digital accumulating scale reported retaining thousands of emails in inboxes and archived folders. Deleting activity is also problematic, with some individuals reporting that they never deleted emails. McKellar, Sillence, Neave & Briggs (2020) interviewed employees about their data management practices and found four underlying drivers of digital hoarding in the workplace: anxiety, compliance, disengagement and collection. While anxiety and disengagement (not enough time/too much effort) resonate with the findings of Sweeten et al. (2018) and Vitale et al. (2018), compliance maybe more specific to the workplace environment in which data is kept (and deleted) in relation to comply with organisation's guidelines and policies.

The McKellar et al. (2020) focused primarily on knowledge-rich organisations i.e., universities, in which academic staff often worked individually in respect to their digital data. In the current study, we expand our understanding of digital data management by focussing on a more commercial organization, where close knit teams work together dealing with commercially sensitive and personal data. We also seek to examine whether technology driven solutions to digital data management are effective and how data management practices are impacted by the workplace setting and culture itself.

We begin by considering the issues associated with personal data, technical solutions and governance before going on to outline the research aims of the current study.

The risks associated with accumulated digital data

Protecting information systems is important given the valuable organisational data resources they hold (Ifinedo, 2009). For any organisation there are risks associated with increasing volumes of stored digital data. Individuals within organisations necessarily hold data, including personal identifiable data. It is difficult for organisations to know the true extent of that data particularly in situations where individuals

are accumulating (either deliberately or inadvertently) emails and other files that may contain a range of different data types, including personal data, for example CV attachments.

Stored material could be mined for social engineering or used by disgruntled employees who have at their disposal a repository of confidential or possibly embarrassing material that may date back many years. Inaction or negligence on the part of an employee may also generate an 'accidental insider threat' in which an individual may inadvertently compromise data integrity. Common examples include leaking sensitive information on social media, mislaying memory sticks, laptops, etc. (Nurse, Buckley, Legg, Goldsmith, Creese, Wright & Whitty, 2014), or simply failing to comply with organisational data protection policies.

The General Data Protection Regulation (GDPR) was introduced in Europe in 2018 (Information Commissioners Office, 2018). This privacy and data protection legislation regulates the storage of personal data. Personal data refers to any information related to a natural person or 'Data Subject', which can be used to directly or indirectly identify the person. It can be anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address. The Regulation places much stronger controls on the processing of 'sensitive' personal data including health information. Personal data and its protection are key features of GDPR. This includes clear legal details regarding its collection, storage, and use limitations. GDPR also requires disclosure of data breaches, for which a high financial penalty may be applied.

Technical and non-technical protection

Organisations have introduced technical solutions to protect against information security incidents and to target data accumulation. These tools and measures include installing firewalls, updating antivirus software, system backups and maintaining and restricting access controls. Technological solutions to reduce data accumulation include the use of General Information Management (GIM) systems i.e., shared repositories, limited email storage or automatic deletion policies whereby emails are automatically deleted after a certain period of time. Technical approaches alone are unlikely to result in protection and therefore greater attention is needed on non-technical approaches – a focus on the employees and the organisational culture. Employees need to follow a number of procedures to counteract security threats. These will vary depending on the organization and its Information Security Policy (ISP). The ISP covers actions employees are expected to take in relation to data management and includes, in relation to data accumulation, information retention policies (IRPs) that govern data storage and deletion. However, the

extent to which employees comply with such polices or circumvent the technical solutions is unknown and there are clear implications for the extent of digital data accumulation within organisations.

We know that information security policy compliance can be a problem for organisations (Bulgurcu, Cavusoglu & Benbasat, 2010; Hearth & Rao, 2009). Drawing on social psychology models of intention and behaviour such as Protection Motivation Theory (PMT) and Theory of Planned Behaviour (TPB), researchers have identified several factors that influence employees' compliance with information security policies (Ifinedo, 2012; Siponen, Mahmood & Pahnila, 2013). These include attitude toward compliance, subjective norms, and response efficacy although factors appear to vary across different behaviours (Blythe, Coventry & Little, 2015). Employees are focused upon the 'day job' and data management is not regarded as a priority (Beautement, Becker, Parkin, Krol & Sasse, 2016; Kirlappos, Beautement & Sasse, 2013).

The usable cybersecurity literature highlights issues with technical and non-technical approaches to data management many of which are likely to resonate with digital data hoarding. However, the specific interplay between policies, organisation guidelines and technical solutions in relation to digital hoarding are underexplored.

Current study

In the current study, we worked with an industrial stakeholder, a large organisation with responsibility for customer and employee personal data in addition to their commercially sensitive data. This provided an opportunity to extend our understanding of digital data management in workplace settings. This study is novel in that it provides employees' perceptions of organisational policies, legalisation and technical approaches to information security and data retention in relation to their digital data management. The context also provides an opportunity to examine digital hoarding behaviours around different data types including personal data.

Focusing on employees' perceptions and behaviours the current study has the following research aims a) to examine motivations and practices around the accumulation and non-deletion of digital data and b) to assess the extent to which organisational attempts to address employee understanding and compliance around data retention are effective.

Method

Participants

Participants were recruited as part of a larger study on digital hoarding in the workplace in which the Digital Behaviours Questionnaire DBQ (Neave et al., 2019) was sent around a large commercial organisation with offices across the UK. This commercial organisation operates a number of different services and has specific policies regarding cloud storage and shared access to files, which we wanted to explore further in terms of digital behaviours.

The DBQ comprises two sections: The Digital Behaviours at Work questionnaire (DBWQ) and the Digital Hoarding Questionnaire (DHQ). The DHQ captures scores on two factors – difficulty discarding and accumulation. High scores are considered over 16 for difficulty discarding and 15 for accumulating, (see appendix for DHQ). In the first phase of that study (N=418) completed the Digital Behaviours Questionnaire (Neave et al., 2019) and at the end of the survey, participants were asked to leave their email address if they were happy to be contacted to take part in a follow up focus group. 145 participants left their email addresses, and of these, 80 participants scored highly on the DHQ. We contacted these high scoring participants, and 11 people (Men=9) were available to take part in the focus group interviews.

In total, three focus groups took place in a quiet location at the participants' workplace. Focus groups were based on location (participants from three different offices from the organisation took part) and participant availability with each group containing between 3-5 participants. Table 1 (below) provides demographic details of participants.

Table 1. Overview of participant's demographic background

Participant	**DHQ Score (Difficulty discarding/ accumulating)**	**Demographic information**	**Job role**
P1	DD:23 A:18	Male, no age given	Technical support
P2	DD:20 A:18	Female, no age given	Security
P3	DD:20 A:16	Male, no age given	Systems engineer
P4	DD:23 A:18	Male, no age given	Project manager
P5	DD:20 A:18	Male, aged 42	Project manager
P6	DD:25 A: 20	Male, aged 30	Security

P7	DD:17 A:20	Female No age given	Security
P8	DD:18 A:16	Male, aged 46	Program director
P9	DD:20 A:16	Male, aged 58	Designer
P10	DD:28 A:24	Male, no age given	Test manager
P11	DD:18 A:20	Male, no age given	Security

Materials and procedure

Following ethical approval from our host institution and the organisation, the DBQ was sent out to staff at a large commercial organisation. Participants were asked to leave their email addresses if they were happy to be contacted to take part in a follow up focus group.

Focus groups were chosen because in our previous work exploring digital behaviours, we found that in interviews settings individuals struggled to decide what was meant by a 'a lot of digital data' whereas in a group setting participants were willing to discuss digital data and come to a common understanding of how much digital data was common (McKellar et al., 2020). Focus groups allowed participants to see that others were behaving in similar ways with respect to digital data and thus they felt comfortable in discussing the topic (Guest, Namey & McKenna, 2017).

The focus group schedule aimed to explore digital hoarding, personal information, GDPR and organisational policies and guidance. Before the focus group, participants were asked to bring along an indication of how many emails they currently had stored in their inbox and an idea of any personal information they might have stored. We used the number of emails that participants had in their inbox as an icebreaker and to prompt discussion among participants regarding data accumulation. The interview schedule covered four key topics: Digital data management, personal information, legislation including GDPR, and policies and guidance. The specific questions are shown below.

Topic: Digital data management

- Roughly how many emails do you currently have in your inbox?
- What is your usual routine for dealing with an email?
- What about attachments in emails – how do you deal with these?
- Do you feel the emails you keep are valuable and useful to you?
- What other types of data do you handle on a daily basis?

Topic: Personal information

- What do you understand by the term personal data?
- What types of personal data do you tend to hold or use in your job?
- Where do you store this data? How do you receive this data?
- Do you believe you have responsibility for personal data?
- Do you believe you have responsibility for sensitive personal data?
- If an email contains personal data, how do you treat it?
- Do you feel that keeping personal data is risky and if so in what way?

Topic: GDPR

- Are you currently aware of your company's policies in relation to GDPR?
- Have you received any workplace training in relation to GDPR? And if so, how did this help your understanding of GDPR
- Do you feel confident at applying GDPR – and how do you manage this in relation to emails?
- Do you feel confident at applying GDPR to other kinds of data you have? (Client data etc if appropriate in spreadsheets/log files?)

Topic: Policies and guidance

- Do you know what your company information retention policy is?
- Do you comply with this for emails and other types of data? If not, why not?
- Do you think current organisations policies make a difference to data accumulation?
- Do you think there are advantages and/or disadvantages to keeping increasing amounts of emails?

The focus groups were facilitated by one member of the research team during August-November 2018, an experienced qualitative researcher with prior experience of conducting focus groups on the topic of digital data management and the average length of the focus groups was 60 minutes. All interviews were digitally recorded and later transcribed verbatim. Participants were informed about the confidentiality procedures in place, how their data was to be used, and that they were free to withdraw from the study at any time without explanation. All participants were provided with an information sheet before the study, signed an informed consent form, and were fully debriefed at the end of the session.

Analysis Procedure

Overall, the aim of our analysis was to provide a clear account of the ways in which digital hoarding behaviours are driven, justified and understood by our participants and how they perceive the role of organisational culture and policy in relation to these behaviours. Guided by Braun & Clarke (2006) we

began by transcribing data from all the focus groups verbatim and the transcripts were then read several times by the first two authors.

We conducted the analysis in two stages. Firstly, we undertook an inductive approach and developed codes that captured different explanations or descriptions around digital data management. The codes were then grouped into subthemes, and then main themes that captured the underlying motivations for hoarding behaviour and the different behaviours around personal data that we saw in the data. This resulted in two main themes: 'Organisational culture versus digital hoarding tendency' and 'Thinking about personal data?'

Secondly, taking a deductive approach, we examined the data to identify instances when participants discussed organisational policies or interventions that impacted on their current digital data practises. We coded these instances according to legislation, information policies and training applicable to the organisation but we were open to identifying new codes within the data. These codes were then grouped into two subthemes and a main theme 'Responsibility versus control'.

Finally, we reviewed the data and the themes with the rest of the research team [blank for review]. During this phase of analysis, any disagreements were resolved through discussion.

Results

Three themes captured the tensions and complexities of thinking about different data types and how they are managed in relation to policies and legislation but also in relation to doing the 'day job'. These themes highlight the potential difficulties for those with high hoarding tendencies in the face of specific organisational processes and information policies. The themes and subthemes are discussed below.

Theme 1: Organisational culture versus digital hoarding tendency

Across the focus groups, there was clear discussion of data accumulation and difficulty discarding. Digital data hoarding was mainly discussed in relation to emails with participants keeping a large number of emails often in excess of 20,000 both in their mailboxes and in email archives. For most participants emails were not organised. Instead, the inbox was simply used as a repository with little attempt at saving useful and deleting less valuable emails. A few participants reported using a folder system to help manage their emails, but most participants used a simple search strategy to locate relevant information as and when needed. Using this strategy was seen as less time consuming than filtering emails or employing a more sophisticated management approach.

No, I, I experimented with a folder structure, but then I got sort of lost of what I had put in what folder, when I was trying to find things, so no I just, I just keep it all in the one inbox, if I do have some folders, for things that are personal, maybe, or none work related or things like this maybe, or something specific, like an event that is going to happen, like, so I might have all the emails for, but like 95% of them are all probably just in my inbox, so I can find everything in one place (P5, FG2)

Although most participants recognised that the vast majority of the emails they kept were not valuable there was still reluctance to delete the data. In part this was in response to the volume of data stored and the time it would take to sort, filter and delete. Participants on the whole thought that this process was simply not worth it although they recognised that it did have the potential to affect productivity.

Yeah, that takes longer. You can search through an inbox, in a fraction of a second to find stuff that you hopefully, find stuff that you want, so, yeah. Even if it is not compliant with policy stuff, it is just easier (P2, FG1)

One of the other disadvantages of having such a large, wodge of email just sitting there is that it is much more difficult to find stuff, if you are trying, if it is not working when you search for key words, and you can't quite find something, so if you did actually spend the time filtering it and getting rid of stuff, that might actually speed things up a bit. (P4, FG1)

Blame culture

The main reason for keeping the emails related to the need for evidence. Participants felt they were often operating within a blame culture. Being able to provide evidence of jobs actioned or certain communications was vital in order to reduce potentially negative outcomes. For these participants,

hoarding data was a natural consequence of the organisational culture in which they worked. All participants discussed the need to be accountable and to be able to evidence work. Some people discussed the threat of being audited both in the external, legal sense and as part of very regular internal processes “the audits that we all get, or are subject to every day” (P11, FG3). This meant participants needed to find a balance between keeping the data longer than they should versus having the data there ‘just in case’ it was needed. Operating in a 'blame’ culture made people anxious about deleting data.

I mean it has been invaluable to me, I mean in the last year, we have used emails for external audits from all sorts of regulatory bodies, and if we hadn’t kept them, then that case would have been, I mean, it would have been really, really bad,... So having that, I mean it is so so valuable having those emails. So I think, I am not sure, because of my role I guess, some of these emails are so valuable for me. I am really reluctant to delete anything at the moment, externally, I mean internal emails, noise, can you do this, can you do that. (P5, FG2)

There will be some query, two or three years down the line that, what was the rationale for going down that route, or you know, why did you do that particular, or did you put that particular bit of code in there, yeah, yeah. You may not remember, you might remember, but if you have got the email audit trail then you know why. (P8, FG3)

Legacy and change

Organisational change also fuelled the tendency to accumulate and retain data. Data acted as a proof not only for work done but also as a proxy for job role or job identity. This was especially important in a situation where individuals held key pieces of knowledge rather than information being stored centrally. Being the ‘last man standing’ meant there was an onus on that individual to keep and provide access to legacy information that would otherwise be lost in a changing environment. This information was useful to speed up work or to help out other people. This was more apparent when participants discussed organisational change and people leaving the company. It was clear that sometimes valuable information about ‘how to get things done’ was not stored centrally but held in emails often over long-term periods to bridge the periods of change.

> *I think from the role I am doing, and what I have been doing for a few years, I deal with a lot of various, individual, projects as such, but there is stuff that I would deal with and I am probably*

the only person that would be able to deal with that particular thing, so I tend to keep a lot of information in emails that, that would relate to a specific topic or a specific answer or I have asked a specific question and something has come back, and I keep a lot of that, I mean I checked and I have emails going back, 2002, which is mad, really. But it is still related to the systems that I use now, and because that system is now legacy and it is underpinned by a lot of the stuff that we are still doing, what we still did back then, I need that information. (P6, FG2)

Theme 2: Thinking about personal data?

Recognising personal data

The majority of participants had a good understanding of what was meant by 'personal data'. Participants often referred to staff training as a way of indicating their knowledge.

I have just done my security training, I should know this. So it is email address, phone numbers, anything, anything that could be described as personal data, so actually every single email in your inbox contains personal data (P9, FG3)

When asked about personal data in the organisation, most participants immediately thought of customer or client data that was held on centralised systems. Here, GDPR was at the forefront of the discussions and focussed on shared access to personal data. Interestingly the participants usually spoke in terms of collective responsibility and used the term 'we' when discussing access and data classification.

Well I had, like I said, that communication around personal data being on our share point, but it was like are we classifying data correctly, are we managing access or reviewing who has access, are we providing different levels of access appropriately. That was, there was a big push around it, that was about a year ago, and I heard that there is a GDPR central compliance team at [organisation] so I have done some of the things, we have kept a list of who has access and we monitor access, we review it and obviously we have a process when someone joins, to make sure that people are removed from having access when they leave. (P5, FG2)

Moving beyond the central storage of customer data, people started to think about where else personal data might be stored and why individual employees might have personal data. Employee data was seen as something that was job role dependent, for example, something that line managers would be

expected to deal with. Participants recognised that this type of personal data could be more distributed and stored in a number of different places and formats.

This is the line managers, this is one source, if you are managing staff then you are going to have more personal data. Because that generates personal sensitive data around the management of people (P2, FG1)

So this is the thing, it doesn't have to be just in an excel spreadsheet, that could easily be in an email. Do you see what I mean? It's in an email exchange I suppose, like that, couldn't it? (P3, FG1)

Participants did not immediately recognise the personal data they had access to or even stored themselves. Overall, people were slow to realise that they even had it. As the discussions broadened, participants began to realise and recall the personal data they had unwittingly received and since kept. Many remarked that they had forgotten about that data though that they probably still had it somewhere.

I find that interesting because I used to manage a team and I still probably got emails from my reports with their personal problems in, you know if they have called in sick or they have had problems with their children or I have had an argument with my husband I can't come in, and it has just occurred to me that, that stuff is still there, and I could probably delete that stuff really, or at least forward it to, you know, whoever the new manager is and then delete it myself, because that is probably stuff that I didn't even want to know in the first place. But, it is still sitting there (P7, FG2)

Processing personal data

The accumulation and retention of personal data followed a different pattern and was accounted for in a different way compared to other forms of digital data (that were more clearly work related). Unlike keeping data just in case (anxiety) the data was stored almost unwittingly and certainly lacked any sense that it was named, organised or could be easily located. Participants did not talk about keeping personal data 'just in case' or for evidence. In fact, participants were more concerned with the difficulties of sharing the data than simply having stored somewhere on their devices/systems.

If that information is given to you by the, by the person, the way I understand it, if that person has trusted you with that, then you have to deal with that data appropriately, and it is your responsibility, but actually having it yourself, isn't a problem, as long as you get rid of it within a certain amount

of time. So I am not sure, if I would be careful with it, I would delete it when I didn't need it anymore, and yeah, it is, it is sharing It and what you do with it that is more important to me (P5, FG2)

Approaches to sharing personal data (usually stored within emails) were characterised by different levels of activity. Firstly, some participants felt that simply not sharing it at all was the safest option. For others, it was important to consider who needed to see this information and to take care when sharing the personal data. Finally, a few participants described the importance of correctly identifying and marking the data as confidential when sharing.

I think in my head, most of my data handling, on [organisation's name] systems around this kind of thing, rolls around, the group of people who will see it, who it will be forwarded to, everything that arrived on the internal [organisation's] system will stay there, it is secure enough for that kind of data, but if it is something personally sensitive about a person, then, just mentally not sending that to people who don't need to know that (P2, FG1)

Participants were aware of the potential risks of having personal data, especially losing personal data. Many referred to the regulations in place to protect data and there was sense that if procedures were followed then individuals would be covered in terms of their liability.

Well you have got the information and handling standard, so the standard tells you how you should, handle that information and we have got, internal information, confidential information and highly confidential information, so sensitive personal data would be classed as highly confidential (P11, FG3)

Theme 3: Responsibility and control

The introduction of technical and procedural changes to promote appropriate data retention posed a number of challenges for participants both in terms of their attitude towards digital data responsibility and in terms of their data management practices.

Personal versus organisational responsibility

Participants appeared to draw a distinction between the limits of their individual responsibility for and control over digital data and that of the organisation. The majority of participants were not overly concerned by the thought that they had personal data stored (perhaps unwittingly) in their inbox. At a

personal level, the risk regarding the storage or hoarding of personal data was seen as low. However, when asked to consider the risks, participants did fall back to the regulations and to GDPR.

It puts you at a risk, if you have it anywhere, like not secure if that makes sense, and also the corporate servers that office 365 are on, it puts them at a huge risk, because that is an even bigger accumulation of data, if everyone used up 50GB of the 100GB then those servers are very, very, lucrative to someone that. (P1, FG1)

Obviously [there is a risk to keeping personal data], if your security was compromised and that [personal] information was stolen, because you are responsible for it (P5, FG2)

However, ultimately many referred to the 'organisation' and its responsibility and its controls. Here, the narrative was around technological security and encryption and the challenges posed by the move to cloud storage rather than on user behaviour per se. The security lens through which participants viewed the protection of data remained squarely with the organisation.

Well, I am assuming now that they have moved over to cloud-based storage, they could pull it all together and have a wholesale policy application, and they can just say, anything, the data is now with us, not on their machines anymore, so we can just do a blanket delete on it. So, I think because I am on that cloud-based storage it kind of takes the responsibility off me and puts it onto them, whoever they are, them, but that is kind of, it gives us more comfort that someone else is worrying about it not me. (P7, FG2)

Technological versus human control

It was clear that tensions existed between the policies and practices deployed by the organisation to ensure compliance (and reduce digital data accumulation) and the day to day working of participants. Participants were aware of some of the technological systems in place to streamline data accumulation either through the use of shared document repositories or through the information retention policy designed to help with the deletion of unnecessary files and data. However, systems that were too prescriptive or processes that removed too much control or took too much time to implement were ignored or bypassed.

The Information Retention Policy (IRP) of the organisation was a document all participants knew about. While this allowed participants to check how long they should be keeping information for, in practice it was circumvented for a number of reasons. Firstly, as described in Theme 1, participants failed to adhere to the policy in order to keep information that they felt would be useful in the future or they retrospectively justified the keeping of data beyond the policy because it had been beneficial for the business.

I think If I am honest, I know probably what we suggest, what it tells us to do, we shouldn't keep it for any longer than we need it, and for the reasons you kept it, but I know fine well I have broken that and I have kept stuff, that I shouldn't really be hanging onto, just in case (P3, FG1)

For others, they simply didn't consider the IRP, it didn't affect their data management behaviours because it wasn't enforced. Participants pointed out that bringing new systems into the organisation created an obvious tension with the IRP. The policy was undermined by seemingly unlimited data storage systems. Unlimited mailbox spaces made it easy to accumulate data.

[retention length] has rather been undermined, by the adding of office 365 into the company, because they previously, they were given you a mailbox of a limited amount of size, so you were forced every now and again to go look at your older stuff, and trim it down or do something with it, but now they have opted to 100 GB so pretty much unlimited. So, the temptation is just to throw it into this large black hole and forget about it (P4, FG1)

Participants explained that the pace of work and the volume of data made it difficult to take the time to review old documents and decide what should be kept and what could be deleted.

I think we are doing so much at pace, that to go back and look at things, that are old and, and decide whether they should be kept, reviewed refreshed would be phenomenal amount of effort, therefore, people just push on, move on and drive forward and just do what we would say they would do, that is two years old, delete it, so it is a lot of effort to consciously to be able to manage everything within a retention policy (P11, FG3).

Central repositories, e.g., SharePoint were also circumvented or not used as intended because of the lack of individual control over the data. Participants felt concerned about using the system because someone else could delete the file or the data. Instead, they would often download a local copy of the

data thus adding to the data accumulation problem. For some participants, they were reluctant to use SharePoint at all and this issue related to management preference and so email copies of SharePoint data was also kept as evidence.

The hardest bit I have got is to actually keep a track of everything that is related to a particular project, because you have got bits that are on email, bits that are in share point, you have got bits that are on teams, instant messenger, on teams and you may or may not have the access rights to all of those areas, so to keep a local copy is, not the ideal solution but it is a way of, at least having your own local audit trail of what, of what change has happened (P8, FG3)

Discussion

This study extends our understanding of digital data management in workplace settings by examining organisation culture and working practices in relation to data accumulation behaviours including personal data management. The study points to four key findings. These novel insights contribute to the small but growing body of knowledge around digital hoarding in the workplace and are discussed below.

Firstly, we note that participants' digital hoarding behaviour was driven by several different motivations. Key amongst these was the need to keep digital data as 'evidence'. Clearly at a broad level this 'anxiety' around data deletion resonates with the work of McKellar et al. (2020) and Sweeten et al. (2018). Participants felt the need to be able to evidence work tasks and communications both in an ad hoc sense and for more formal audit processes. The sense of anxiety in relation to digital data was underpinned by a perception of 'blame' culture within the organisation which highlighted to participants the potential damage to both individuals and the organisation of not being able to provide evidence through digital data.

However, examining the context around the 'anxiety' expressed in this study provides a more nuanced account of this motivation that differs from previous work. In earlier research 'anxiety' or a sense of keeping data 'just in case' often captured participants' general rather than specific unease at deletion behaviour. In those instances, participants were often unable to express the value of the data or offer any specific example of how or why the data might become useful in the future. In the current study, all the participants articulated very precise and detailed examples of how and why the data they'd kept had become useful, often invaluable. Participants used these examples to justify their reluctance to delete data given the organisational culture in which they worked. Their 'anxiety' was clearly driven by the 'blame

culture' within the organisation rather than as a response to individual differences as evidenced in some of the previous literature. Importantly, this primary motivation was seen across all participants regardless of status from technical support to project manager. The consensus around this driver indicates that anxiety in relation to organisational culture needs to be considered more structurally in relation to the systems and processes in place rather than perceived as purely an intrinsic or personality issue.

We also saw evidence of participants' behaviour around digital data being driven by response to organisational change. Here, keeping digital data prevented knowledge from being lost (Holten, Hancock, Persson, Hansen & Høgh, 2016) during times of staff turnover and organisational restructuring or system change and provided a sense of security and identity for some of our participants. Emails reflected important legacy knowledge (Dineen & Krtalić, 2020) and were seen primarily as part of the organisation's collective knowledge rather than as individual legacy.

Secondly, regulation and local policy did not always drive digital data hoarding. In previous work, compliance with local policy was seen as an underlying motivation for digital hoarding. However, McKellar et al. (2020) did suggest that compliance may be a factor that is more sensitive to the organisational setting than other motivations. We find support for that suggestion in this study. Although our participants were clearly aware of policies and guidelines regarding data management, including the Information Retention Policy (IRP) and GDPR, and the majority were able to articulate the organisation's stance concerning data management more generally compliance with the regulations and policies did not drive data practises. On the contrary, at an individual level, we found participants often chose not to comply with the IRP. Non personal data was kept at an organisational level in order to comply with, for example, external regulatory audit purposes but at an individual level, many participants did not actively comply with IRP but kept local business data for knowledge hoarding purposes or to provide evidence for local internal checks. In fact, many reported using workarounds both for personal and organisational reasons. People felt that despite the policies it was in their best interests and ultimately those of the organisation to keep data beyond the guidance specified in the Information Retention Policy (IRP).

Thirdly, personal data is thought of and dealt with differently in terms of digital data management practises. The McKellar et al. (2020) study did not address participants' perceptions and behaviours concerning personal data directly. In this study we note that participants were less likely to keep personal data 'just in case' but were more likely to have simply forgotten they had it or felt that keeping it stored was less risky than sharing the data. In this sense, personal data did not appear to be stored for 'legitimate purposes' but was held unwittingly. There was almost a sense of ambivalence towards the personal data

they held and a feeling that it was relatively harmless while it was left undisturbed. At the same time, participants recognised the GDPR practices surrounding organisational held personal data. These apparent contradictions reflect the notion that privacy is contextually bounded, and that social norms and practices shape the ways people think about personal data (Park & Shin, 2020).

Finally, we found that technical solutions and systems to centralise repositories and manage data can result in a tension between organisations and employees and can actually be counterproductive where digital hoarding is concerned. Participants highlighted the inconsistencies between policy and practice. For example, requirements to delete emails when storage limits were reached being overridden by new systems providing seemingly infinite data storage for emails. The use of systems implemented to reduce additional digital data storage e.g., centralised data repositories was also undermined by specific local practices and preferences and by concerns over control and access. The blame culture effect was apparent here with participants often downloading additional copies of emails and other documents to ensure they personally had the 'evidence' they needed.

The remainder of the discussion focuses on possible technical and non-technical approaches to reducing these tensions and improving digital data management. Finally, we revisit the potential threat posed by digital hoarding to organisations and end on suggestions for future research.

Organisational factors and non-technical approaches

Organisations, through their procedures and policies, may unwittingly drive the hoarding behaviour of their employees. The apparent tension between technological systems and data management plans can lead to employee confusion and reduced compliance with policies. In large organisations, both the organisation and its employees are slow to change their practices. Although training can increase awareness of data stored, in particular personal data, it is not sufficient on its own to change behaviour (Branley-Bell, Coventry, Sillence, Magalini, Mari, Magkanaraki, & Kalliopi,2020; Michie, Atkins & West, 2014). Management and organisational factors affect the cybersecurity culture of the organisation (Da Veiga, Astakhova, Botha, & Herselman, 2020) and tackling employee attitudes, norms, personal agency and habit can also influence employee behaviour (Coventry & Branley, 2018; Glaspie & Karwowski, 2017). Individual differences, such as hoarding tendency or people's management of their own privacy in employment settings vary (Neave et al., 2019; Hargittai & Litt, 2013) but signalling the importance of data privacy and better data management behaviours more broadly could be incorporated into the organisation's practices. With respect to digital hoarding, one approach may be for organisations to

include data management as part of their employees' performance appraisal. Giving data management more of a focus will encourage feedback and attention from management and signal its importance within the workplace. Providing this focus could allow employees to feel they have permission to spend time engaging in data management activities. Employees often fail to comply with security behaviours in the workplace because they prioritise the 'day job' and feel the response cost of compliance is too high. Individuals and organizations place different values on the cost and benefits of different security policy behaviors – the so-called compliance budget (Beautement, Sasse & Wonham, 2008) although it is apparent that there are different costs associated with carrying out different behaviours (Blythe et al., 2015) and this may impact employees' behaviour.

Technology centred approaches to digital data hoarding

Technology can help streamline and support user actions around digital data retention and deletion. Email management tools work by providing users with reminders and snoozing functions promoting action at some point in the future (Gwizdka, 2001, 2002). Many of these functions could be automated and some researchers have suggested that platform automation could enact these kinds of personal information management (PIM) actions for less tech savvy people (Alon & Nachmias, 2020). While this may be fruitful in a broader PIM setting, the constraints of a workplace setting make automated processes more difficult to manage for both business and individual reasons especially in situations where individuals are happy with their PIM practices as was the case in our study. Tools that make it easier for users to think more carefully about the data they have stored also exist (Vitale et al. 2018). Although, for these tools to be effective they need to match individuals' needs and preferences with regards to personal data management (Vitale, Odom & McGrenere, 2019).

In a workplace setting in which personal data is a key component, it will be important to consider how the issues raised in our study concerning locating and processing personal data could be supported through technical solutions. PIM studies have investigated different ways of improving file search and retrieval with some researchers advocating a user subjective approach to PIM with systems developed around attributes of the files such as project, importance and context (Bergman, Beyth-Marom & Nachmias (2003). While these suggestions relate to improved productivity, they have clear implications for the storage and processing of personal data and policy compliance in this space as well.

While technical solutions have predominantly focused on individuals, organisations have a role to play and could take steps to reduce their mass emails and consider more carefully who exactly needs to see

the information. Organisational level technical solutions designed to centralise repositories for security and data reduction purposes are widespread particularly across larger organisations but can have unforeseen consequences in relation to digital data hoarding. Group information management can have more usability problems than PIM in terms of data organisation and retrieval (Bergman, Israeli, & Whittaker, 2020) leading to reduced uptake. Systems are not always well accepted if existing solutions or workarounds are already in place and well used (Ammari, You & Robert Jr, 2018). Participants are likely to make comparisons between any new system designed to help manage data and those systems already in use. Researchers note that task, team and culture factors as well as the presence of existing technologies are important factors in the uptake of new systems and employees can often end up using 'all the systems' to varying degrees rather than simply using the new system as a direct replacement (Ammari, You & Robert Jr, 2018). Although usually considered in relation to workflow, further research on group collaborative technologies and shared repositories is needed from a digital data management perspective. In sum, organisational changes (both technical and people based) can lead to counterproductive work behaviours and it may be that unpopular changes are more likely to trigger further hoarding.

Organizational risks relating to digital data management behaviours

Beyond reduced productivity, digital data hoarding has the potential to cause harm to organisations through insider threat. Insider threat comes in two forms. Firstly, from individuals who act out of malice and use their insider knowledge to intentionally compromise an organisation's data security, for financial or personal gain, or for revenge for an actual or perceived slight. The second type of insider threat is accidental and can arise from simply failing to comply with organisational data protection policies. Both types have equally serious consequences for an organisation, and since the introduction of GDPR in May 2018, the fines for data breaches can be significant.

Digital hoarders may represent a significant risk to their employers as they are, firstly, less likely to distinguish between valuable and mundane information – simply storing it all in an unsecured fashion – and, secondly, more likely to have damaging information available, should their loyalty to a company be challenged, i.e., should they become an insider threat. Research has suggested that certain aspects of personality are predictive of workplace compliance behaviours. For example, individuals high in agreeableness are more worried about what other people think about them and are more concerned with workplace security issues (Korzaan & Boswell, 2008; Shropshire, Warkentin, Johnston & Schmidt, 2015).

In determining the 'threat' posed by digital hoarders in this context exploring other aspects of personality such as narcissism may be an important next step (Spain, Harms & LeBreton, 2014).

Limitations and Future work

In terms of future research, this study has suggested a number of areas and issues that could be explored further. Firstly, while participants recognised personal data, they often questioned its sensitivity in terms of its potential value to others. It may be that further explorations of information sensitivity are valuable in this context (Blythe et al., 2015). Second, further work exploring the tensions and perceived mismatches between policy and local practice may prove fruitful in identifying key points at which employees decide to ignore the IRP. It is important to understand more about the organisations 'norms' that are perhaps being unwittingly signalled to employees. Alongside an audit of procedures and practices, a clearer of understanding of local practices around group technologies should include a clear reference to digital data hoarding. Third, future studies should look to develop interventions that take account of the underlying motivations or drivers of data hoarding. Identifying when interventions are more likely to have an effect is also important. Employees can only accumulate large amounts of data if they have been with the organisation for a longer period. Those new to the organisation will have had less time to accumulate data. Previous research has shown that those in management positions exhibit greater levels of digital disorganisation than more junior colleagues (Uğur & Çalışkan, 2022). While we did not explicitly collect data on how long the participants had work for the organisation, it was clear that none of the employees were newly in post. It would be useful to speak to new starters as well as to individuals with considerable experience in different organisations. This would help establish triggers for the development of good habits and a better understanding of what worked well in other organisations and why. It is also worth considering that only people happy to talk about their data management behaviour volunteered to take part in our study and that others identified as scoring highly on the DHQ might have different views and behaviours.

Finally, focusing on employees and raising awareness about the data they hold is an important part of the story and a key first step to personal agency. In terms of enacting change, it will be important to include multiple perspectives including management perspectives as a more holistic approach is required. Line managers may be useful in terms of providing a richer perspective, and this is something we are going to investigate going forward. This study has identified that the values and culture of an organisation are an

important factor in the attitudes and behaviours around digital data management and as such future work will need to explore whether the findings reported here are seen across different large-scale organisations.

Conclusion

This study examined the digital data management practices of employees working in a large commercial organization. The participants described their hoarding behaviours predominantly in relation to the working culture of the organization and detailed the tension between technical solutions designed to streamline data accumulation and loss of control and local data management preferences. While technical solutions will undoubtedly have a place in future data management strategies, organizational norms and practices will have to be considered in parallel if such solutions are to be successful. Tools to support digital data management need to further consider the type of data that is being kept as well as the drivers of data accumulation in order to be effective.

References

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

Ammari, T., You, S. & Robert Jr L.P (2018) Alternative Group Technologies and Their Influence on Group Technology Acceptance. *American Journal of Information Systems*, 6, 2: 29-37. Doi: 10.12691/ajis-6-2-1.

Beautement, A., Sasse, M. A., & Wonham, M. (2008). The compliance budget: managing security behaviour in organisations. *In Proceedings of the 2008 New Security Paradigms Workshop* (pp. 47-58).

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).

Bergman, O., Israeli, T., & Whittaker, S. (2020). The scalability of different file-sharing methods. *Journal of the Association for Information Science and Technology,* 71,12: 1424-1438.

Bergman, O., Beyth-Marom, R., & Nachmias, R. (2003). The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology*, *54*(9), 872-878.

Blythe, J. M., Coventry, L., & Little, L. (2015). Unpacking security policy compliance: The motivators and barriers of employees' security behaviors. *In Eleventh Symposium on Usable Privacy and Security* (pp. 103-122).

Branley-Bell, D., Coventry, L., Sillence, E., Magalini, S., Mari, P., Magkanaraki, A. & Kalliopi, A. (2020). Your hospital needs you: Eliciting positive cybersecurity behaviours from healthcare staff using the AIDE approach. Annals of Disaster Risk Sciences.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology,* 3,2: 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*, 34,3: 523-548.

Coventry, L. & Branley, D. (2018). Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *Maturitas* 113, 48–52. DOI: https://doi.org/10.1016/j.maturitas.2018.04.008

Da Veiga, A., Astakhova, L.V., Botha, A & Herselman, M. (2020). Defining organisational information security culture—Perspectives from academia and industry. *Computers & Security*, 92, DOI: https://doi.org/10.1016/j.cose.2020.101713

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

Glaspie, H. W., & Karwowski, W. (2017, July). Human factors in information security culture: A literature review. In International Conference on Applied Human Factors and Ergonomics (pp. 269-280). Springer, Cham.

Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field methods*, 29, 1: 3-22.

Gwizdka, J. (2001). Supporting Prospective Information in Email. In CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01). ACM, New York, NY, USA, 135–136. DOI: http://dx.doi.org/10.1145/634067.634150

Gwizdka, J. (2002). Reinventing the Inbox: Supporting the Management of Pending Tasks in Email. In CHI ’02 Extended Abstracts on Human Factors in Computing Systems (CHI EA ’02).

Hargittai, E., & Litt, E. (2013). New strategies for employment? internet skills and online privacy practices during people's job search. IEEE security & privacy, 11(3), 38-45.

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

Holten, A.-L., Robert Hancock, G., Persson, R., Marie Hansen, Å. and Høgh, A. (2016). Knowledge hoarding: antecedent or consequent of negative acts? The mediating role of trust and justice. Journal of Knowledge Management, Vol. 20 No. 2, pp. 215-229. https://doi.org/10.1108/JKM-06-2015-0222

Ifinedo, P. (2009). Information technology security management concerns in global financial services institutions: is national culture a differentiator? Information Management & Computer Security, 17 (5) pp. 372-387

Ifinedo, P. (2012). Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. Computers & Security, 31(1), 83-95.

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

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

Korzaan, M.L., & Boswell, K.T. (2008). The influence of personality traits and information privacy concerns on behavioural intentions. Journal of Computer Information Systems, 48: 15-24.

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

McKellar, K., Sillence, E., Neave, N., & Briggs, P. (2020). There is more than one type of hoarder : collecting, managing and hoarding digital data in the workplace. Interacting with Computers, 32(1), 209-220.

Michie, S., Atkins, L. & West, R. (2014). The Behaviour Change Wheel : A Guide to Designing Interventions. Silverback Publishing, London, UK.

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

Nurse, J.R.C., Buckley, O., Legg, P.A., Goldsmith, M., Creese, S., Wright, G.R.T., & Whitty, M. (2014). Understanding insider threat: a framework for characterising attacks. IEEE Security and Privacy Workshops, DOI: 10.1109/SPW.2014.38

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

Oravec, J. A. (2018). Digital (or virtual) hoarding: emerging implications of digital hoarding for computing, psychology, and organization science. International Journal of Computers in Clinical Practice (IJCCP), 3(1), 27-39.

Park, Y. J., & Shin, D. D. (2020). Contextualizing privacy on health-related use of information technology. Computers in Human Behavior, 105, 106204.

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

Shropshire, J., Warkentin, M., Johnston, A., & Schmidt, M. (2006). Personality and IT security: an application of the five-factor model. AMCIS Proceedings, 415.

Siponen, M., Mahmood, M., & Pahnila, S. (2013). Employees' adherence to information security policies: An exploratory field study. Information & Management. 51. 10.1016/j.im.2013.08.006.

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

Spain, S.M., Harms, P., & LeBreton, J.M. (2014). The dark side of personality at work. Journal of Organizational Behavior, 35 (S1): S41-S60.

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

Uğur, N. G., & Çalışkan, K. (2022). Time for De-cluttering: Digital Clutter Scaling for Individuals and Enterprises. Computers & Security, 102751.

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

Vitale, F., Odom, W., & McGrenere, J. (2019, June). Keeping and Discarding Personal Data: Exploring a Design Space. In Proceedings of the 2019 on Designing Interactive Systems Conference (pp. 1463-1477).

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

Appendix

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

	Not 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