The paper first appeared in the Journal of the Association for Information Science and Technology on June 15, 2015. http://dx.doi.org/10.1002/asi.23526s

Towards understanding short-term personal information preservation: a study of backup strategies of end users

Matjaž Kljun[1], John Mariani[2] and Alan Dix[3]

[1]Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Slovenia
[2]School of computing and communication, Lancaster University, UK
[3]School of Computer Science, University of Birmingham, UK

October 30, 2014

Abstract

The segment of companies providing storage services and hardware for end users and small businesses has been growing in the past few years. Cloud storage, personal network-attached storage (NAS) and external hard drives are more affordable as ever before and one would think that backing up personal digital information is a straight forward process nowadays. Despite this, small group studies and corporate surveys show the opposite. In this paper we present the results from a quantitative and qualitative survey of 319 participants about how they backup their personal computers and restore personal information in case of computer failures. The results show that the majority of users do manual, selective and non-continuous backups, rely on a set of planned and unplanned backups (as a consequence of other activities), have inadequate knowledge about possible solutions and implications of using known solutions, etc. The study also revealed that around a fifth of all computers are not backed up, and a quarter of most important files and a third of most important folders at the time of the survey could not be (fully) restored in the event of computer failure. Based on results several implications for practice and research are presented.

Keywords: information preservation, information backup, computer backup, personal information management, information space maintenance

1 Introduction

In the digital information age the preservation of digital information is a challenging issue. Based on Berman's Data Pyramid [4] there are three levels on which digital information needs to be addressed. At the top is information valuable to whole society usually managed by governments, national libraries or health services. The middle has information important to communities, educational and research institutes or companies. At the bottom lies the personal information of individuals that is valuable to one or maybe a few people (e.g. family photos). Moving up the pyramid the risk of loosing data decreases while responsibility and stability increase.

Management of information in higher levels of the pyramid is entrusted to professional technicians, information managers and librarians, and the budget for setting up and running the preservation infrastructure is usually planned for. The preservation of information at the bottom of the pyramid is left to the individuals who might not have adequate knowledge of available technologies and solutions. The long-term information preservation has been described as challenging for individuals (obsolete (proprietary) file formats, software, and storage technologies). However, even short-term information preservation is an issue as there are plenty of information loss stories in circulation. Nonetheless, short-term preservation is very important to fulfil the different needs and obligations of an individual. In this paper we present the results from a qualitative and quantitative study of a short-term information preservation and backup practices of 319 end users and their personal views on the subject.

2 Previous work

Information management of an individual is at the core of the research area called Personal Information Management (PIM) which studies activities people perform in order to acquire (keep), organize, maintain and retrieve ((re-)find) personal information such as files, web bookmarks and email messages. The maintenance of personal information is described in PIM frameworks [2, 15] as a process of deciding on composition and preservation of information including archiving, deleting, and backing up information. However, it has been noted that maintenance often has the lowest priority of PIM activities [2, 6, 14]. Time spent on maintenance for future (unknown) benefits is hard to justify with today's busy schedules and it is often done sporadically (e.g. moving files from the desktop when it becomes cluttered) or during major life changes (e.g. changing job).

Several PIM studies observed backing up as a side observation. One study revealed that managers often rely on system administrators for backup services and have no knowledge about how and how often this happens [3]. However, the ways they manage their information could result in its loss. Nevertheless, with cheap storage, fairly reliable systems, and older information of low value, backup and maintenance will be non-issues for these managers. Kaye et. al. reported that very few participants in their study were backing up personal information and the ones who did were motivated by their own bad experience or by the stories of their peers [18]. Backups were mostly improvised, unplanned and inventive activities involving tools primarily used for other things (e.g. email [12]). Even if users claimed that they could get some version of their lost documents from the duplicates (e.g. sending a file to a colleague) the findings show that disk failures without appropriate backups have negative effects on productivity in a similar proportion to major physical changes such as office moves, for example.

Several studies reported on a variety of information preservation techniques [14, 17]: (i) on a network drive, (ii) manually copying files to external media, (iii) a source control system, (iv) synchronising information between two computers, (v) emailing documents to themselves, (vi) burning information on optical drives, and (vii) duplicating files within the same computer. This variety of strategies shows that people find different ways of preserving important information. However, not all of these techniques can be counted as a backup and just a minority of users used an automatic procedure. In our own study about how people manage fragmented information participants used several complementing backup strategies [CITE] greatly affected by ICT knowledge – many participants doing tedious manual copying were not familiar with other available options. Even if not backing up they claimed to be able to restore information to a certain degree (printed documents, shared files, public repositories, etc.) Another study reported that information fragmentation makes some processes such as backup harder to perform [30]. While documents (usually stored in a home folder) are easy to backup, this is harder for bookmarks and email (if not accessed over IMAP or web) which are stored in folders hidden from users.

Besides the above small studies where backup was only a side observation, Marshall devoted several studies to long-term preservation of digital belongings [24, 25]. Her findings revealed that the majority of users approach preservation with benign neglect. The ad-hoc storing of digital belongings according to perceived future use, exigencies of the moment, and as a side effect of other tasks results in distributed information across devices and various web services (sharing photos, videos, emailing documents) which leads users to many false thoughts. For example, users think that in the long run they can keep track of and remember everything that is important to them, and that tools and techniques to access this information will be available infinitely. However, even in the short run the majority of participants failed to implement simple backup, had a hard time finding backed up information, or have lost information online (deactivated accounts, services bought out, hacked, etc.) Her suggestions to long-term preservation involves changes on the part of the users as well as service providers and include: (i) designating and assessing value of information to be curated, (ii) creating a central catalogue of distributed repositories, (iii) automating some services (e.g. virus checking, keep formats up to date), and (iv) accessing information in new modes (e.g. visualisation).

Commercial companies providing backup services or hardware also regularly conduct information preservation surveys. A report from 2012 summarised answers from 2,209 individuals [19]. Only 3% of respondents backed up more than once a day, 7% once a day, 20% at least once a week and 36% at least once a month (presumably people used more than one strategy with different frequencies, which the study does not reveal). A high proportion of users (43%) rarely or never backed up their information. Another report from the same year [32] claimed that only 10% of the respondents backup continuously, 30% daily, 30% weekly, 30% monthly and 25% never back up. Of the ones who back up, 27% use an external hard drive, 18% USB thumb drive, 9% burn CDs or DVDs and 8% use cloud storage. Symantec's research has been targeting corporate users [34] and reports that organisations (as individuals) increasingly use a blend of different backup strategies which complicates backup and recovery of information. There were several incentives reported why companies backup their information (business continuity, disaster recovery, compliance, security, and redundancy) [8]; however, the study revealed that about a quarter of companies never check the reliability of their backup strategies.

It is very clear from both academia and corporation studies that a significant proportion of computer users are not concerned about doing backups. Although backups cannot be extrapolated to work over decades (removable media deterioration or loss, file formats and storage obsolescence, services' closure) the very thought of preserving information in the short run is nonetheless a first and important step towards long-term preservation. Little is known about the motivations behind why users do or do not backup, if they do what they actually backup, what are the strategies and how they complement each other, and how users restore information in the event of storage failure. This study aims at revealing more detailed explanations behind (not) backing up motivations and strategies of preserving information in the long run.

The paper is structured as follows: next two sections present the motivation and methodology, followed by the results sections (divided in backup strategies, information restoration and backup stories). The paper finishes with discussion (implications for practice and research) and conclusion.

3 Motivation for study and research questions

This study was motivated by an earlier study about how people manage their fragmented information [CITE]. As in other studies, backups were a side observation. When asked if participants would be able to restore important information they mentioned various restoration sources that we combined into two groups based on whether they were initially intended for backup or not.

Planned backup
A planned backup contains a copy of personal files and folders used primarily for preservation of information. Of 19 participants only five used a backup software (two run it manually) and two relied on a corporate backup. Others were manually copying documents to an external hard drive, printing files, copying information in the cloud, and sending files as email attachments. These procedures occurred quite irregularly and were selective on which information to include.

Unplanned backup
This occurred when users had sent documents to others via email, had a copy on a thumb drive for transferring between computers, the work was published and thus publicly available, documents were printed out for reading, were shared in an online project repository or participants simply relied on other people to possess their documents.

The unplanned backup was often considered as a restoration point as there was a great chance that it contained a newer version. The manual and infrequent backups were the main reason behind such inconsistencies; users showed a general distrust in automatic procedures, the need of control over what has been backed up, and the lack of knowledge about available technologies. Many admitted the bugs in their strategies but were stuck with old habits and the immediate benefits of backing up were vague which affected the motivation for improving them.

In an effort to understand more about how users try to preserve the information we disseminated a questionnaire focused on:

(i) How users backup their computers?

- What are the main backup strategies and how do these relate to background (age, gender, knowledge, years using computer) and other variables (type of computer, operating system)?
- How frequently do users backup their computers and the relation to background and other variables?

- What is the relation between manual, automatic or semi-automatic backup procedures and background and other variables, and what triggers non-automatic procedures?
- How much of the different file groups (personal files, software profiles/preferences, software) do users back up and does this relates to background and other variables?
- What and why users backup or why they do not?

(ii) The use of planned and unplanned backups as restoration points

- Are users able to restore current important documents and folders and to what extent?
- How and where from would these documents/folders be restored and is the restoration source a part of planned or unplanned backup strategy?

4 Method

For this study we created an online questionnaire. The questionnaire was improved through 4 iterations with 30 volunteers with different computer knowledge recruited through convenience sampling. For each iteration we used a different group of participants. After each iteration we gathered opinions and suggestions for improvements, and looked at the responses to see if answers were in the context of each question.

4.1 Description of the questionnaire

The questionnaire was divided into 4 parts. The first part included general questions about age, gender, subjective knowledge (on a 1-10 rating scale), and number of years using the computer for creating documents. Other questions in this part were number of computers used, type of computer (personal or company and laptop or desktop and other), operating system and whether the computer has RAID. The next two parts of the questionnaire were based on our research questions.

The second part was about backups of personal computers mentioned in part one. It was repeated maximum twice (even if more computers were mentioned) so as not to make the questionnaire too long. The questions were presented in a matrix where along the top were backup strategies: external hard drive, CDs or DVDs, cloud, email, network drive and other. For each selected strategy a frequency, type (manual, semi-automatic, automatic), and percentage of three groups of files (personal files, software preferences or profiles and software) backed up by that strategy had to be selected. Besides percentage for each group of files an open-ended question for additional explanations about what files of that group are backed up needed to be answered. If no strategy was used, an open-ended question to explain why that computer is not backed up needed to be answered.

The next part of the questionnaire inquired about the three most important documents and folders respondents worked on or used in the last weeks and whether they would be able to restore them if lost (yes, no or partially) and where from. If no or partially were selected, an open-ended question asking why was needed to be answered. Respondents were also asked if the restoration source was primarily intended as a backup or not.

The last part involved concluding questions about how respondents estimated their happiness with their backup strategies (on a 1-10 rating scale), and two open questions about why they assigned a particular rate and if they have any backup stories to share which might have changed their backup strategy planning in the past.

4.2 Dissemination, participants, and validity of estimates

We used a combination of convenience and snowball sampling starting with our own contact lists and social networking contacts. Besides asking users to circulate the questionnaire among their friends and colleagues, we advertised it among students at our institutions and on a few general purpose web forums. It was sent to mailing lists of at least 10 universities and posted on social networks of at least 20 individuals. The questionnaire circulated in several European countries besides Japan, Australia, United states, Mexico, and Canada from February to April 2013. It was answered by 319 individuals who took

The paper first appeared in the Journal of the Association for Information Science and Technology on June 15, 2015. http://dx.doi.org/10.1002/asi.23526s

on average 30 minutes to complete it. The shortest was 10 minutes while one participant spent 3 hours to finish it (we received an email from her telling us that she knew nothing about backups and how important they are until seeing the questionnaire).

Participants' (Table 1) age ranged from 17 to 70; 80% of them have been using computers (for reading, writing and creating content) for more than 10 years. Males tend to use more computers than females (two-sample t-test t(274.595) = 3.33, p-value < 0.001) and rated their knowledge higher on a 1-10 rating scale (two-sample t-test t(187.475) = 6.57, p-value < 0.001).

Table 1: The nature of respondents

	All respondents		**Females**		**Males**	
N	319		106		213	
	mean	**s.d.**	**mean**	**s.d.**	**mean**	**s.d.**
Age	31.66	10.21	31.36	9.77	31.82	10.44
Years using computers	15.48	5.9	14.28	5.86	15.82	5.9
Computers used	2.2	1.25	1.9	0.98	2.34	1.34
Subjective knowledge rate	7.55	1.98	6.55	2.01	8.06	1.76

The estimation about subjective computer knowledge was compared to whether respondents know what RAID is by an indirect question *"Do you use disk mirroring (RAID) on this computer?"* with possible answers *yes*, *no* and *don't know*. There was a significant difference in the rated subjective computer knowledge between respondents who knew what RAID is (mean = 8.04, s.d. = 1.79) and those who did not (mean = 6.16, s.d. = 1.83) based on the two-sample t-test (t(138.721) = 8.05, p-value < 0.001). This suggests that knowledge about what RAID is, does have an effect on the estimated value of one's computer knowledge and we are assuming that estimations of subjective computer knowledge are adequate.

Respondents using Windows XP rated their subjective knowledge lower (mean = 7.1, s.d. = 1.8) than those using Windows Vista and 7 (mean = 7.5, s.d. = 1.9), Windows 8 (mean = 8.7, s.d. = 1.7), OS X (mean = 8.1, s.d. = 1.8) and Linux (mean = 8.1, s.d. = 2). A one-way analysis of variance revealed significant differences between the groups (F(4,537) = 5.02, p < 0.001). Other background variables such as age (linear model, p-value: 0.6467) and number of years using computer (linear model, p-value: 0.1083) were not influential on the computer knowledge.

Estimates of the percentage of files being backed up by different strategies had to be selected on a percentage scale ranging from 0 to 100 with a 10% steps in between. As the accuracy between estimates and actual amounts cannot be faultless (based on preliminary observations) we will look at the results in 4 ranges: 0%, more or less than 50% and 100%.

Last estimation – happiness about backup strategies on a 1-10 rating scale – is used only to catch the general subjective happiness towards information preservation and there is no need to validate it.

5 Results: Backup strategies of personal computers

The following subsections are base on the research questions presented in Section 3.

5.1 Backup strategies and their combinations

Participants described backup strategies for 542 computers (319 main and 223 secondary computers). We considered the main backup strategy the one that is done most frequently, has the highest percentage of files backed up, its storage size is bigger than computers' internal hard drive, and based on the open-ended answers. The summary can be seen in Table 2: 30 (9.4%) main computers and 81 (36.3%) secondary computers are not backed up (111 or 20.5% of computers). By far most popular strategy is backing up to an external hard drive followed by a cloud storage and network drive.

When divided by computer type (Table 3), company desktops are mostly backed up on a network drive while other computer types are backed up on hard drives. Laptop computers have higher usage of cloud storage than desktop computers. Copying to CDs and DVDs is rarely used as a main strategy,

Table 2: Main backup strategy by main and secondary computer

	External hard drive	CDs or DVDs	Cloud storage	Email	Network drive	Other	No backup
1st Computer	137 42.9%	3 0.9%	76 23.8%	31 9.7%	27 8.5%	15 4.7%	30 9.4%
2nd Computer	62 27.8%	2 0.9%	36 16.1%	9 4.0%	23 10.3%	10 4.5%	81 36.3%
Both	199 36.7%	5 0.9%	112 20.7%	40 7.4%	50 9.2%	25 4.6%	111 20.5%

but rather as complementary one (used for 76 computers with a frequency of every few weeks – see the right side of Figure 1). The main strategy for 40 (7.4%) computers is just emailing files. Other main strategies include backing up to another internal hard drive (13), personal cloud (2), personal server (6) and NAS (4).

Table 3: Main backup strategy by type of computer

	External hard drive	CDs or DVDs	Cloud storage	Email	Network drive	Other	No backup
Personal desktops	61 35.9%	3 1.8%	30 18.0%	10 6.0%	5 3.0%	12 6.6%	48 28.7%
Personal laptops	99 44.8%	4 1.8%	53 24.0%	19 8.6%	7 3.2%	6 2.7%	33 14.9%
Company desktops	24 24%	0	17 17%	2 2%	31 31%	4 4%	22 22%
Company laptops	21 37.5%	0	14 25.0%	2 3.6%	8 14.3%	3 5.4%	8 14.3%

Table 4 shows combinations of backup strategies – on average, each computer is being backed up by 3. Combinations of external hard drive, cloud storage and/or email are most common. 27 computers are backed up solely in the cloud and 19 computers by other strategies. This is followed by combinations of a hard drive, CDs or DVDs, email and/or cloud storage. The remaining combinations include less than 10 computers.

Table 4: Combinations of backup strategies (ones that occurred 5 or less times are not listed).

# of occurrences	83	42	36	35	28	27	19	18	16	12	12	9	9	8	7	7	7	6	6
External hard drive	●	●	●	●				●	●	●	●	●		●	●	●			
CDs or DVDs								●	●	●	●								
Cloud storage		●		●		●				●	●		●		●	●		●	●
Email		●	●					●		●			●	●			●		
Network drive					●							●		●	●			●	
Other							●									●			●

There is no significant relation between the main backup strategy and years using computers or gender. However, it is affected by subjective knowledge. Email as a main backup strategy falls with higher subjective knowledge: it is used by 66% of those who rated it 2, it falls down to around 10% for those who rated it 5-7, and down to 4% for those who rated their knowledge 10. Cloud storage is mostly used by respondents who rated their knowledge 7 and falls down with higher (some of these respondents were concerned about the privacy as discussed later) or lower knowledge rate (participants who did not know about such solution). The relation between the main backup strategy for each computer and operating system is shown on the right side of Figure 2 and will be looked at in the next section.

5.2 Frequency of backing up

The frequency values available for each selected strategy were: all the time, several times a day, once a day, every few days, once a week, every few weeks, every few months, other and don't know. The distribution of frequencies by backup strategies can be seen in Figure 1. On the left side are frequencies for main strategies used on each computer being backed up. A high proportion of low frequencies, especially with external hard drives as the most common backup strategy, can be noted. The majority of computers using cloud storage, network drive and other are being backed up daily.

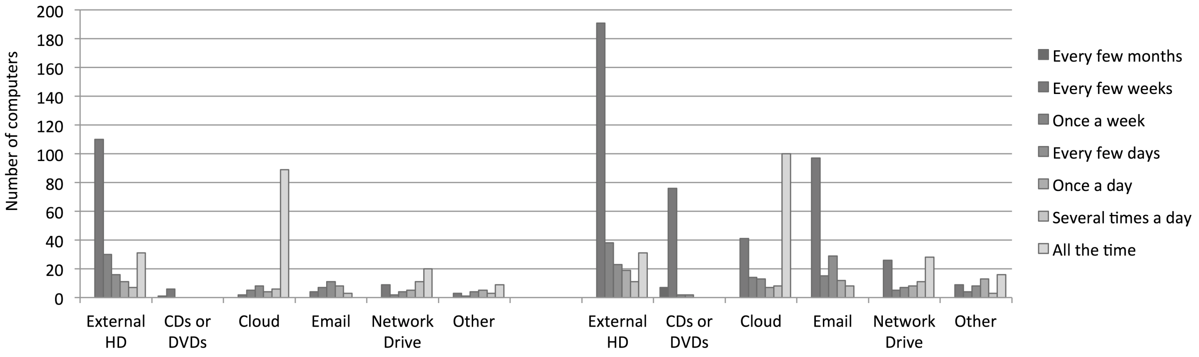

Figure 1: Frequencies of initiating backup for different backup strategies. Left: frequencies for each main backup strategy. Right: frequencies for all strategies used (multiple per computer).

On the right side are frequencies for all strategies (multiple strategies per computer). It is apparent that the secondary and further backup strategies tend to be less frequent. External hard drive, email and CDs or DVDs are most popular secondary strategies. Out of 542 computers only 155 (28.6%) are backed up all the time (31 on a hard drive, 100 in the cloud, 28 on the network drive and 16 other).

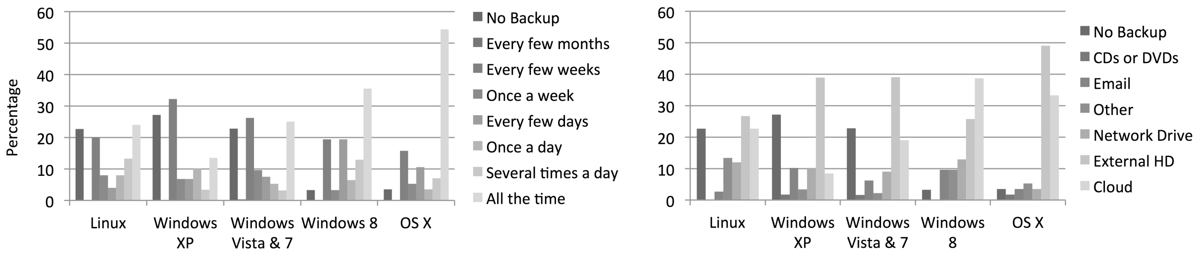

Figure 2: Left: Frequencies of doing backup by operating system. Right: Main backup strategy by operating system. Windows operating systems are grouped based on the integrated backup solution.

One-way analysis of variance has shown a significant effect of years using computer (divided into 5 years intervals from 0 to 30) on frequency of backups (F(5,350) = 3.57, p = 0.0036). Similarly subjective knowledge has significant effect on frequency density (F(8,365) = 2.16, p = 0.03). In both cases Bartlett's test did not show violation of homogeneity of variances ((5) = 2.89, p-value = 0.72 and (8) = 2.35, p-value = 0.97). This suggests that computer knowledge is acquired through usage and that valuable information accumulates over years together with awareness of possible loss. Looking at gender, more males then females were not backing up their computers (10.3% vs. 5.7%); although males were using more frequent backup procedures than females – e.g 34.7% males doing all the time backup vs. 21.7% females.

Company desktops and laptops are backed up more frequently (daily backed up 54% and 50% respectively) than personal desktops and laptops (daily 31.7% and 34.4% respectively) as seen in Table 5. This is not surprising as there is usually an IT service available in companies. Laptop computers are backed up more (around 85%) than company and personal desktop computers (78% and 71.3% respectively). This might be because people are aware that laptops are more prone to physical damage (answers to open-ended questions often mentioned fear of drop, theft and precautionary backing up before or during travels).

More than half of OS X based computers are backed up all the time on hard drives as the main strategy (see Figure 2). The reason is probably a popular continuous data protection software Time machine integrated in the operating system (nearly all of these computers use it) since 2007. Windows 8 has a similar solution called File History introduced in 2013. In this study only a handful of users were using it and we cannot compare it to Time Machine. Backup solutions in other Windows versions (Backup and Restore) offer only scheduled backups which was reported in Microsoft's own statement as underused [27] and only a few users used a third party solution.

Table 5: Frequency of backing up percentage by computers type.

	Personal		Company	
	Laptops	Desktops	Laptops	Desktops
N (542)	221	165	56	100
All the time	**26.2**	**23.4**	**28.6**	**36**
Several times a day	4.1	3.0	12.5	8
Once a day	4.1	5.4	8.9	10
Every few days	9.0	9.6	5.4	4
Once a week	11.3	6.6	5.4	6
Every few weeks	**29.9**	**23.4**	**25.0**	**14**
Every few months	0.5	0.0	0.0	0
Not backed up	**14.9**	**28.7**	**14.3**	**22**

5.3 Manual, automatic or semi-automatic backup

Only 160 (37.1%) off 431 backed up computers have a fully automated backup procedures in use, 62 (14.4%) a semi-automated (where backup is mostly automatic but some user intervention is needed to initialise it) and 209 (48.5%) use manual procedures (everything done by the user).

Automatic procedures gradually drop (the right graph on Figure 3) as the frequency of backups gets tenuous (from 6% to 71%) while manual procedures rise in the same direction (from 6% to 83%). The question is why computers with automatic procedures are not backed up more often. The reason stated were that backing up slows down computers (*"Backups are a nuisance. They take time to set up, generally require some form of maintenance, and when a scheduled backup is running it uses a significant chunk of your computers resources (slowing down everything, and make YouTube videos jumpy for example)."*) and when prompted users do not plug in external drives.

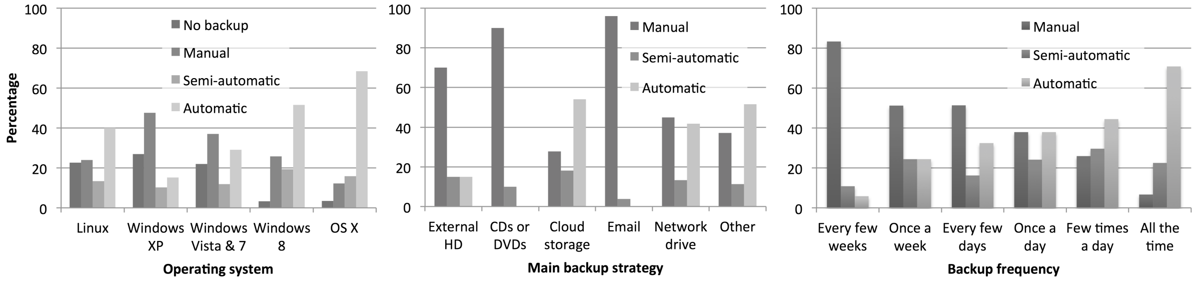

Figure 3: Manual/automatic backup procedure by *Left:* operating system (Windows operating systems are grouped based on the integrated backup solution), *Centre:* main backup strategy and *Right:* backup frequency.

The centre graph on Figure 3 shows the manual/automatic backup procedures divided by the main backup strategy. An interesting observation is the use of manual and semi-automatic procedures in combination with cloud storage: several users treat a could storage dedicated folder in a similar fashion as external hard drive and manually copy (duplicate) files from their hierarchy to a cloud storage folder. Users feel as they do not have ownership and control over these folders, they do no fit in the organisational scheme, and might behave differently than other folders. This distrust can be extracted from answers to open-end questions such as ‘‘*I have my own folders which don't fit in Dropbox ... I copy only a few files there*”, “*I deleted a Dropbox folder on one computer and it got deleted on all of them when I just wanted to disconnect the first one*”, “*It is stored [information] in God-knows-which internal folder that Microsoft [Sharepoint] manages*”, ‘‘*This system might crash or be unavailable so I might lose the access to the latest versions of documents [edited on other computers] ... So I always like to have local copies.*”

Users of Windows 8 (51%), Linux (40%) and OS X (69%) use more automatic procedures than users of Windows XP (15%) and Windows Vista or 7 (29%) as is seen on the left graph on Figure 3. This also correlates with rated subjective knowledge of users of each operating system. The percentage of people using an automated procedure is dropping with lower subjective knowledge: 48.4% of those who rated their knowledge 10, 31% who rated it 9 or 8, and around 20% who rated it 7-4. The number of people not doing any backup is dropping in a similar way the other way around: 40% of those who rated their knowledge 4, 27% of those who rated it 5, 40% of those with 6, 35% with 7, 26% with 8, 23% with 9 and 20% of those who rated it 10. There is also a statistically significant (linear model, $p < 0.001$) effect of orated subjective knowledge on type of backup (automatic vs. semi-automatic vs. manual); the automatic procedures are more likely used by people with higher subjective knowledge.

The company computers were in the forefront in applying automatic procedures as well (automated procedures can be of different frequencies): 44% of company computers use automatic backup procedure compared to 35.3% of personal laptops and 24% of personal desktops. This means that strikingly 56% of company and around 70% of personal computers still require some user intervention. One of our study aims was to uncover what triggers the manual or semi-automatic backup process. Below are the summaries for each backup strategy.

External hard drive

The frequency of manually backing up files greatly depends on the “feeling” about the possibility to lose information when time and effort are invested in it, when it cannot be recreated (e.g. backing up photos when downloading them from a camera), at milestones (approaching due dates, finished project stages), when internal drive runs out of space, when installing or upgrading to a new operating system, when computers get slower/louder, before travelling, after experiencing or hearing about information loss, and when computer shows a reminder. In the latter situation the backup is often postponed and forgotten as users are in the middle of another task.

CDs or DVDs

Users reported to burn one or more copies of disks when they are running out of space, before reinstalling the operating system, after milestones (e.g. finished project, end of year), after holidays when photos are uploaded from cameras. Other less frequent triggers were burning disks for others (unplanned backup) and a “feeling” that something might go wrong. Some use CDs and DVDs for special file types only: photos, audio and video files or software installations (licences).

Cloud storage

The cloud storage folder is most often used for the work in progress. The ones who treated this folder as an external hard drive are copying files over either regularly (e.g. end of the working day), on special occasions (e.g. before going on or during a trip) or simply when they remember. Others use the cloud storage as a temporary working space and documents are moved out after the work is finished. Cloud storage was often used for other purposes rather than backup (access, sharing) which resulted in unplanned backup. While some respondents reported to store and back up only the most valuable documents (e.g. tax reports, visas, files with passwords) others were concerned over the trust issues (data mining).

Email

The use of email as a backup solution has been already mentioned in other studies [12, 22, 26]. This study revealed several other rationale behind this behaviour. Email is the most common strategy of unplanned backups (e.g. sharing and transferring between computers). Email as a backup strategy is a quick, dirty and ubiquitous solution with no additional software involved. The attached files can be accessed from “anywhere” (respondents mentioned network restrictions at work). Several users used email to backup documents during travels. A few respondents used it solely as a backup strategy and emailed all new and edited documents after a day of work, project's end, after a deadline or after every major change in documents. It has been also described as more error-prone and secure than external or thumb drives. Three major advantages over the cloud were reported: email is opened and checked all the time, works as a reminder, and provides always visible dates as a proof of postage.

5.4 Backup of different file groups (personal documents, profiles, software)

The average percentage of files backed up by the main strategy is 69.3% of personal documents, 28.5% of profiles (or 47.5% on computers that do have profiles backed up) and 18% of software (or 50.3% on those who do back up software).

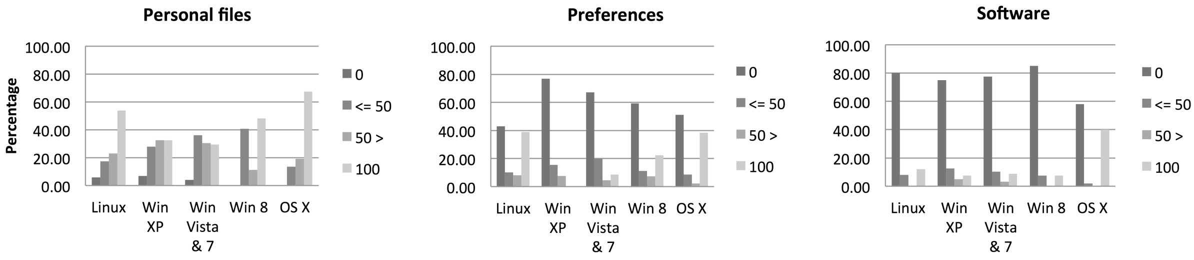

Figure 4: Types of files (personal, profiles/preferences, software) being backed up by the main backup strategy by operating systems (Windows operating systems are grouped based on the integrated backup solution).

We divided the percentage estimates into 4 groups: 0%, less or equal than 50%, from 51 to 99%, and 100% (see Figure 4). As with automatic procedures and frequency, the OS X based computers had the highest percentage of personal files, software profiles and software backed up. This indicates again that a good pre-installed backup solution plays a role in the amount and types of information backed up. which is supported also by open-ended answers: *"Mac is fully backed-up, Win is not. OS X Time Machine does it automatically, in Win I just backup the user folder."* or '*'Time Machine creates a complete mirror of the primary and secondary disks."* Windows 8 based computers have a higher percentage of profiles backed up than Windows XP, Vista and 7. The move of software profiles from a dedicated folder in XP to a users' home folder (as in Linux and OS X) might be the reason for this. Another explanation is that Windows 8 users might also be more advanced as early adopters and based on subjective knowledge.

Several people reported no backup solution for files; however, they backed up web browsers' profile in the cloud. The reasons stated were the simple integrated browsers' solutions. Some profiles are backed up to be able to quickly restore work environments in the case of a computer failure or replacement. The respondents who selectively backup profiles are relatively happy with the default settings for the majority of other installed software. Several users mentioned preserving all profiles as they backup their home folder. Respondents who do not backup profiles claimed either (i) that profiles do not contain any important information, (ii) do not know how to do it or (iii) where to find them.

Of 31 computers that are fully backed up, 17 use OS X, 6 Linux, 6 Windows 7 and 2 Windows 8. Of these 22 on a hard drive, 7 network or NAS, 2 in the cloud; 23 are backed up automatically and the rest semi-automatically; 12 are backed up all the time, 5 few times a day, 4 once a day, 2 few times a week, 4 once a week and 4 every few weeks. All but one fully-backed up computers with OS X use a pre-installed backup solution. Among 34 computers with Windows 8, only 4 (11.8%) were using pre-installed software, compared to 14 of 313 (4.5%) computers with Windows Vista and 7.

People backing up 100% of personal files are most of the time less then 40 years old. There is a significant change in backing up higher percentages of files in this age group. When running a general linear model between three age groups (15-19, 20-40, 40+) we found that each of the age groups has a statistically significant ($p < 0.001$) effect on backing up 100% of personal files compared to other age groups. One explanation might be that people over 40 began to use computers later in their lives (which corresponds with the frequently mentioned lack of familiarity with technology). However, the age group below 20 also does not backup a lot. When coding their answers the cited reasons were not many important files to lose and mostly easily recreated or retrievable at that stage of life. As before, this suggest that knowledge (when users start to use computers) and awareness of possible loss of slowly accumulated information play a role in backup strategies.

People who rated their knowledge higher also tend to backup more profiles and software (linear model, $p < 0.001$). Gender wise, 41% of males do full backup of personal documents (compared to 34.8% females), 20% do full backups of profiles (9.5% females), and 14.3% full software backup (against 11.22% females). Percentage of files is also related to a backup strategy. For personal files, hard drive is by far the most popular solution for backing up 100% of the files, followed by network, NAS or personal server. The majority of computers with less than 50% of personal files being backed up use the cloud storage solution – mainly the basic free account with limited storage capacity. Backing up 100% of personal files does also not guarantee the higher frequency – only 24.8% of those who back up all personal files do it all the time (34.2% every few weeks). The ones who fully backup all 3 groups of information 38.7% do it all the time. We discussed earlier the reasons behind these lower frequencies.

5.5 What and why users backup or why they do not?

We have already mentioned a few reasons behind backup decisions in previous sections (e.g. use of the cloud as external hard drive or no accumulation of important information for people under 20). Here we will look at what is being backed up by each strategy and information group.

Personal Files

External hard drives and network drives are used for backing up everything. If the size is of concern, documents, personal photos and videos, and other information that cannot be easily restored, found or recreated has a preference. If information is easily accessible elsewhere (e.g. publicly available music, videos, software installations) some respondents would not waste backup storage, except if such files are a part of an often-accessed large collections built over time (e.g. legal documents, scientific papers). *CDs and DVDs* are used to archive important documents, personal photos and videos. Several participants complained that their collections of CDs/DVDs is becoming hard to catalogue. Respondents not concerned about privacy/security use *Cloud* for important information currently in active use, documents that need to be shared or accessed from different devices, documents that would take too much time to recreate, and files of which past versions are important. The space limit of basic cloud storage accounts is the reason why more information is not stored here. Some users use dedicated services for dedicated file types (music, photos and videos). *Email* is also used for currently worked on documents, documents that need to be accessed from different devices, photos and URLs. As with the cloud storage, size, convenience, privacy and safety have surfaced in answers. Users mentioned also that email for backup used less than in the past.

Profiles

If not backing up everything *external hard drives* are used to backup web browsers' profiles (or just parts: bookmarks, passwords and certificates), email clients' profiles, settings for wireless networks, game profiles, profiles of remote session connections, IDE profiles, business software profiles, designers' software profiles and catalogues, colour profiles for monitors and printers, etc. to support quick recreation in case of need. *CDs or DVDs* are used only by a handful of participants to backup their email archives. General *cloud storage services* were mentioned to be used only for parts of web browsers' and email clients' profiles (bookmarks, passwords, email accounts). Otherwise dedicated cloud storages for synchronising profiles of web browsers, games and similar software that features cloud synchronisation was used. Some users used a dedicated cloud storage for backing up entire computers. *Email* was mentioned to be used to store passwords and bookmarks and *other storage* to store roaming user profiles.

Software

Respondents who selectively backup software usually backup installation files of bought software (together with licences), special software not available online, software too large to be quickly downloaded and self written software on either *external hard drives* or *CDs/DVDs*. Several respondents backup drivers so they have them ready when re-installing their operating system. Some mentioned doing an image of their hard drive occasionally on another hard drive.

Several reasons why computers are not backed up have been brought to light. As mentioned before, many younger respondents do not have (enough) sensitive and/or important information that would be worth spending time and money for setting up a backup strategy. Quite a few also reported not having any bad experience and thus do not need backup. Some claimed that RAID mirroring is enough and some do not have additional storage to do it. Several of them think that backing up is too difficult, they have no knowledge of how to do it, or others are doing it for them. And nonetheless two claimed they did not know about backing up and its importance.

6 Results: Restoration of important digital information

Respondents were asked to name the three currently important files and three folders, if they would be able to restore them if lost, and where from. Of 957 files, 5.4% would be partially restored while 18.6% would not be restored. Of 957 folders, 12.7% would be partially restored while 17.8% would not be restored. In Figure 5 we see the percentage of files (left) and folders (right) that would be either restored or not for people using different main backup strategies. Respondents using cloud as their main strategy would restore most of the documents while people using CDs and DVDs least documents. The percentage of partially restored folders is higher compared to files which was expected since unplanned backup (e.g. sharing information via email) often occurs with files rather than whole folders. Even respondents who do not do any backup would be able to fully restore 60% of files and half of the folders. This shows a big reliance on unplanned backups.

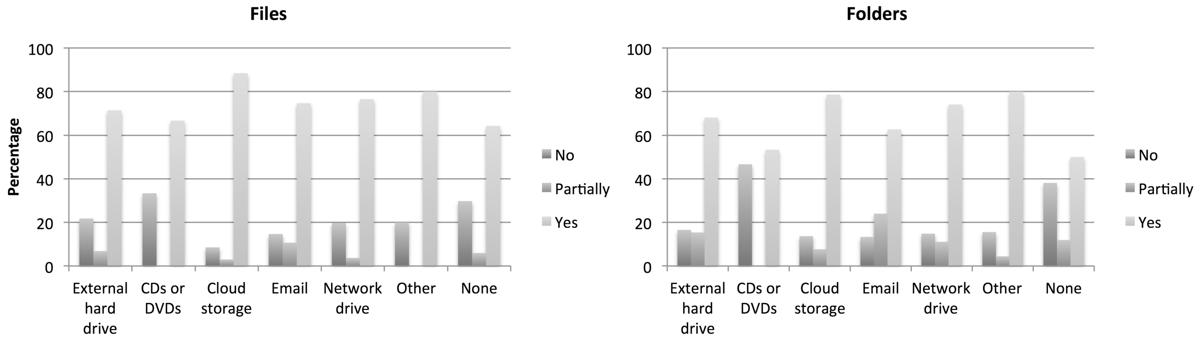

Figure 5: Percentages of no/partially/fully restored files and folders divided by the main strategy used.

Answers to where would files and folders be restored from often included more than one restoration source. Of all files 61.9% would be fully and 58% partially restored from the main backup strategy storage. If hard drive, cloud and other were the main strategies 70-75% of files would be restored from a planned backup, if CDs and DVDs only 10%, if email 38% and if network drive 20%. The rest of files would be restored from either other backup strategies or unplanned backups.

Of all folders 66.1% would be fully and 71.5% of partially restored from the main backup. If external hard drive was the main strategy, 84.9% of folders would be restored, if cloud 66.6%, if other 63.9%, if CDs/DVDs 25%, and if network drive or email 21%. The frequency of backing up on CDs and DVDs is usually low and documents exist in other unplanned backups. Unplanned backups as restoration point was claimed to be faster as it eliminates red tape in companies.

So far we came across multiple references to the (unplanned) backup as a consequence of other PIM activities such as: "*It's not really a backup. I just store photos and music on an external drive for convenience.*", "*I copy on a thumb drive to work on another computer and meanwhile make backup.*", "*Email is more for accessing files from different computers than for a backup.*" or "*When I need to access it from multiple places or want to collaborate it goes onto Dropbox. It is not part of a systematic backup scheme.*".

In this part respondents were directly asked if restoration source was primareily meant for backup (see Table 7). About a quarter of all files (and less folders) would be restored from sources not meant as a backup and only around half of all folders and files would be restored from sources whose only role is a backup. This shows that respondents are very aware of the mix of unplanned and planned backup strategies. However, the fact remains that nearly a quarter of the most important files and a third of the most important folders would not be restored to the current state (the numbers are probably higher

Table 6: Percentage of fully and partly restored files and folders by planned and unplanned backup strategies.

	Files		Folders	
Source	Partially restored	Fully restored	Partially restored	Fully restored
Other uses besides backup	21.15	27.10	26.23	21.80
Only for backup	51.93	46.35	51.64	56.54
Not for backup	26.92	26.55	22.13	21.66

for information not deemed important). The main two reasons are either not doing backups or low frequency between them. Rationale for the former were: not enough time and effort invested in the files or folders, could probably be retrieved from someone or the internet, the size of files or folders too big to fit on available backup storage, information was not shared with anyone yet or will be irrelevant soon, an external drive has stopped working, or respondents have run out of empty CDs/DVDs. The rationale for the low backup frequency were: lack of knowledge about how to backup automatically, files or folders have not yet been sorted or organised to be backed up, respondents forgot to backup or simply could or did not take time yet to start a backup process. And as mentioned earlier the reminders often interrupt users at inconvenient times.

The mixing of restoration possibilities shows the variety, complexity, and fragmentation of planned and unplanned backup combinations – documents exist in many versions, instances and copies.

7 Results: Backup stories

The last question was about backup stories they would like to share and that might have changed the way respondents back up in the present. Not surprisingly, they were mostly about data loss or about difficulties involved in backing up.

We could divide the reasons behind data loss into 4 groups: (1) the lack of backup, (2) not understanding how technology works, (3) technology failure and (4) making mistakes while managing information and computers. The majority of users who experienced data loss claim to make redundant backups or backup more regularly. Interesting enough, five respondents reported loosing information twice before investing in a backup solution and quite a few still do not regularly backup despite experiencing data loss. Participants still regret loosing information over a decade ago even if they do not always remember what exactly got lost.

Restoring information is a daunting process if there is no backup or if backup fails. Although there is an explicit distinction in respondents descriptions between files that can be recreated (e.g. written material) with only time being at stake, files than can be found again (e.g. online) and files that can not be restored at all (e.g. photos).

A big problem is understanding how technology works: respondents claimed they accidentally deleted important files from a versioning systems and did not know how to restore them, they opened attached files directly from emails (saved in some temporary folder and edited) and were later unable to find them, did not know that CDs/DVDs get corrupted, they claimed having antivirus/RAID/thumb drive does not require them to backup, or did not know that most cloud storage services do versioning. People also find it difficult to backup because they are not aware of the possibility to automate the process: *"Backing up all the time is too much work, because then I loose track of what is old and what is new material. Very difficult to organise it all for me"* or *"There's a lot of things to backup and you better have a todo list with things to backup in order not to forget any of them."*

Nonetheless there are also many (users' or technology) accidents resulting in data loss: formatting the wrong drive, not checking the validity of a backup before reinstalling the system, accidentally deleting files, pouring liquid over a computer, lightnings, technology failures (external and internal hard drives, optical disc burners not burning properly, backup software creating corrupted copies), when manually copying from one drive to another overriding new with old files, installing OS on a wrong drive, or backup software did unexpected actions (e.g. a merging function deleted the original files when expected to leave them and just copy new files from a merging folder).

8 Discussion

Although we tried to reach a broad spectrum of users, we understand that participants interested in the subject were more likely to participate. For example only 52 (16.3%) respondents rated their subjective knowledge 5 or less which might explain why the average percentage of backed up computers in our study is higher than in studies presented in section 2. There is also a selection bias in age of respondents as most of them are aged between 20 and 40. Nevertheless, based on the number of respondents and their descriptions of backup related problems we believe that the study captured a general picture of the digital information preservation behaviours.

For the majority of respondents backing up is not an easy task. The mean happiness on a 1-10 scale was 6.73 (s.d. = 2.55) and its relation to subjective knowledge is statistically significant (linear model, $p < 0.001$). Many reasons for being unsatisfied surfaced throughout the questionnaire several times: non systematic error-prone manual procedures (low frequency based on remembering to backup, accidentally replacing files, not all files being backed up, manual versioning), backups are not remote, to expensive to improve, lack of time or knowledge to improve it, not being sure if backup works correctly, not trusting backup outsourcing (e.g. cloud), backup technology is not reliable, backup is taking users' time, consuming computer and network resources, solutions are limited to a certain location (e.g. home), automatic procedures hindering sense of control, and backup software too complex to understand or configure properly.

Frequency of doing backups (two-way ANOVA, $F(5,368) = 4.6368$, $p < 0.001$), automated procedures (two-way ANOVA, $F(2,372) = 10.753$, $p < 0.001$) and the percentage (divided in 4 groups) of personal files backed up on main computer (two-way ANOVA, $F(3,371) = 13.832$, $p < 0.001$; see right graph on figure 6) have a statistically significant effect on happiness. As can be seen from the left side of Figure 6, respondents who use external hard drive and cloud storage are happier than ones who use email and network drive as their main strategy.

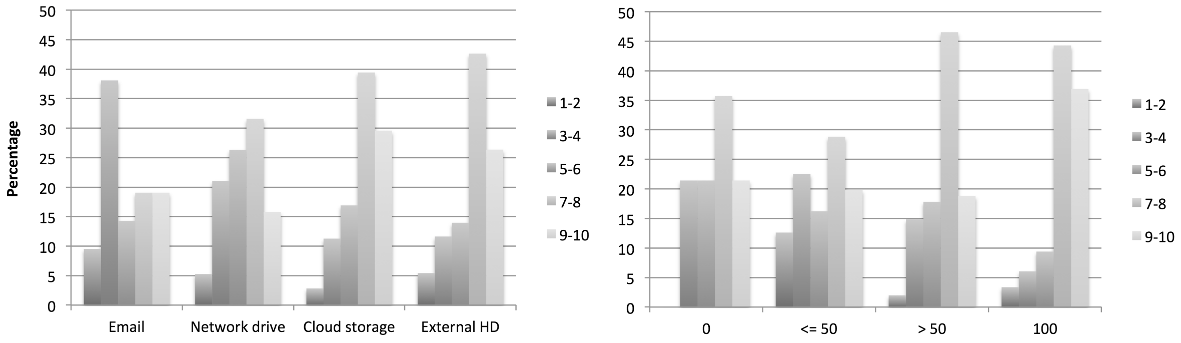

Figure 6: Estimation of happiness with backup strategy on a 1-10 rating scale vs. *Left:* four mostly used backup strategies and *Right:* percentage of personal files backed up with the main strategy.

The results of this study have several implications presented in next subsections.

8.1 Implications for practice

8.1.1 Backing up

The backup plans of the majority of respondents are inadequate for short and long term information preservation (e.g. partial backups, no redundancy, no remote backups, etc.) The continuous data protection with the main backup strategy is used on only 22.1% of all listed computers and 20.4% of computers are not backed up at all. The rest need human intervention to commence the backup process. Nevertheless, in the shade of other information activities (sharing, synchronising, accessing from various computers, etc.) a lot of personal information is backed up in an unplanned way. As such, backup procedures are an interwoven set of planned and unplanned backups which consequently introduces several disadvantages (backup fragmentation, manual versioning, metadata loss). A lot can be improved in promoting the importance of digital preservation and technologies for doing it.

Awareness of backup importance and knowledge of available technologies

In general, awareness of backup importance, possible solutions, and possible consequences of using them is not high. Several respondents revealed a false sense of security with claims that (i) their hard drives are safe enough, (ii) thumb drives cannot fail, (iii) mirroring hard drive is a backup plan, etc. There is also misunderstanding between archiving (e.g. moving unwanted information from internal to external media to free up internal space) and backing up (duplicating information) as apparently several users archived information (see Section 5.1).

About 10% of respondents never heard of cloud storage and many do not know implications of using it. Some users mentioned the risks of data security, confidentiality and availability. However, many users are not aware of network security, data locality, integrity, segregation, access, breaches, authentication and authorization, web application security, virtualisation vulnerability ... [33].

Respondents noted that activities such as manually moving files, versioning and selective backups do not require additional software and provide some sense of control (*"I feel safe backing up personal information to something I can hold and be responsible for its physical security."*); nevertheless they also noted that such activities are highly prone to errors. Several respondents used such procedures because of not knowing any better.

There is a need to educate users about what is possible to preserve, in what ways, and what are possible consequences of each scenario. Prior to choosing the technology, users should first decide on what they want to backup (personal files, entire computer), how often and how many old versions need to be preserved, encryption, etc. Only then the technology that would address their need comes into focus.

Better/easier to use/intuitive backup software solutions

An issue that was mentioned several times in our study is the amount of computer resources needed to run a backup software which calls for light-on-resources and less intrusive solutions. Another one was the lack of understanding backup software as some users mentioned the loss of information from backups because of the interpretation of the interface (for example one user reported the need to merge two folders which led to a non-recoverable removing of the destination folder). There is a need to better user interfaces and standardised terminology. Another problem (in terms of PIM) are cloud storage services interfering with current hierarchy organisations forcing users to use a predefined folder only. While there exist services offering backing up different parts of file hierarchy, users opted for most popular solutions.

Backup software should be facile to use, non-intrusive and always in sight. For example computers with pre-installed continuous data protection backup software that is functional almost out-of-the box proved to be more backed up than computers without such an option. The developers of backup software must understand that backing up personal files is one of those activities users do not want to be bothered with until things go wrong. Nevertheless, there must be a balance between control and automation. Extremely adaptive systems can change the system in unexpected ways, thus impeding the understanding and prediction of past and future changes. The sense of being in control and the level of challenge perceived in using computers were characterised as very important to the flow of human-computer interaction [13].

8.1.2 Restoring

The ultimate goal of a backup is to be able to restore information. As reported nearly a quarter of businesses do not test their disaster recovery plans [8] and the numbers are probably even lower for personal users. It has been suggested that the plethora backup strategies can have a negative effect on information restoration [24]. We have observed, on the contrary, that when backup strategies are not systematic and frequent, the variety of backup is clearly favourable in the short run (even users with sporadic or no backup affirmed to be able to restore a significant amount of important information). However, while having a greater chance of restoring information, the distributed and replicated information bring disadvantages in terms of versions' curation and loss of metadata (e.g. a creation time) [23]. Such reliance on unplanned backups and infrequent backup procedures work against the long-term preservation and clearly need to be readdresses by users.

8.1.3 Long term backup

Although this study focused on a short term storage, backing up is a first step towards long-term digital information preservation. The changes in storage technology is not a concern as it only requires moving files from old to new media [1] and as in short-term backup technology failures can be addressed with information replication. The file (often proprietary) formats are more problematic since nobody can assure accessibility of related software in the long run. There have been numerous calls for open formats [1] to preservation, accessibility and compliance with legislature. While some attempts have been embraced by the industry (e.g. PDF/A for digital documents) other proposed standards have not been adopted (e.g. Everplay for consumer images). It is even harder to preserve web bookmarks and desktop email clients profiles (availability of related software), let alone web based services.

Jones divided information irreplaceableness into four groups: (1) precious irreplaceable information such as photos, (2) extremely difficult to replace information such as tax records, (3) reference collections such as personal academic papers repository, and (4) working documents. He claims that all groups need a short-term backup, whilst only the first two groups need long-term preservation [15, p 160]. Our observations have shown that respondents are aware of vulnerability of and take more proactive stewardship towards preserving some information (e.g. digital photos) over the other (e.g. music, movies or even working documents). Marshall notes that selectivity in backing up information (culling) as observed in our study might be rooted in limits of human attention (less information more aware are we of it), misunderstanding of technology (removing files from hard drive to speed up the PC), and desire of having control over the information space (removing unknown files) [26]. However, our belief is that manual procedures triggered by "gut feelings" based on current preservation needs are the background of selectiveness.

Some researchers also advocate a service type of personal digital preservation [15, 26, p 160]. However, if we entrust our information to such services, there is no guaranteeing that they will still be available in the years to come (e.g. Ma.gnolia or Megaupload). For example, one of the authors has lost 3 years worth of records with the now defunct Posterous blog service as he had used it to syndicate the content to other services. Even if Posterous let him backup or download all his posts, it was of no use since the syndicated links now all point to a non-existing domain. Nonetheless, the size of our personal digital archives might also not be easily uploaded and downloaded as the rate of the archives grows faster that the rate of the internet speed connections. We are not saying that this might not change, but for many in the developed world, DSL internet connection or size caps are still the reality.

In a digital world, a life-long preservation faces many challenges which were almost non-existent in the physical world, whereas it also provides new opportunities to perform tasks that were not straightforward before (e.g. having multiple off-site copies as one of the respondents: *"Every month I send 10 nicest photos to my mother and mother in law. If something happens I will still have copies."*). Despite this, long term preservation is facing many challenges and needs to be reconsidered at various levels. And many short term backup strategies (e.g. storing photos on CDs) do not provide sufficient long-term support.

8.2 Implications for research

There are two areas of personal information preservation not yet covered in the PIM literature. One is how individuals value different kinds of information needed to be preserved in the light of different preservation needs, and the other is how are they motivated in acquisition of knowledge and adoption of new IT solutions and technologies to actualise preservation goals and needs.

Value of personal information

In order to initiate motivation (in our case to preserve information), it has to be preceded by values and needs [21]. According to Maslow's hierarchy of needs there are six layers of needs that a person satisfies in a linear fashion: psychological, safety, love, esteem and self-actualisation. Information preservation is part of a safety level providing security for (among others) resources and property. It has been shown numerous times (including in this study) that when personal resources are endangered (e.g. valuable information on a failed hard disk) users have a higher awareness of potential loss and higher motivation to preserve them. It has also been observed that thinking of a potential loss decreases the confidence level which leads to fulfilling the lower safety needs (e.g. from a respondent's comment *"Unfortunately this [information backup] is not a well organised part of my life and I need to do it ASAP."*)

For the purpose of personal information preservation, its value can be studied and observed from two points of view: economic and personal. An economic value might be applied to an information item if we (or company we work for) sell or charge for services (e.g. proofreading based on number of pages) or if we are legally bound to hold information (e.g. tax records) and there are a fines involved if we fail in doing so. A value can be explained by several theories: (i) cost-of-production (value as a sum of resources needed to achieve a state of an information item), (ii) intrinsic (value of an information item that is contained in an item itself regardless of preferences of an individual), (iii) subjective (value, from a perspective of a buyer derived from subjective concious decision based on feelings, desires, intuitions – not facts of reality – to fulfil needs [35, p. 94]) or (iv) objective value (value to a specific individual, for a specific reason in a specific context based on evaluation of known facts and relation to other values [29]).

In the terms of economic value, respondents most frequently mentioned the time and effort invested (cost-oh-production) in an particular information item or collection. Thus the value of information lies in the amount and quality of the labour required to produce it as perceived by its creator at a certain moment in time (regardless if this labour was completed for personal or economic reasons). Objective theory of value says that the value constantly changes based on all facts an individual applies in valuing information in a particular context. This means that besides invested resources other (objective) factors can affect the value of an information item as well (e.g. a due-date for a short and simple document). However, time, effort or other tangible external factors often play no crucial role in valuing certain kinds of information (e.g. family photos). Rather, the value is determined subjectively and is based on our whims or emotions (see e.g. [28]) and other related reasons (e.g. a wish to preserve information for descendants). We divided different kinds of information by incentives and different periods for preservation in Table ??.

Table 7: Kinds of information based on irreplaceableness divided by incentive and different periods for preservation.

		Preservation	
	Short term	**Middle term**	**Long term**
Economic incentive for preservation (significant resources invested in creation)	Working documents (project documents & collections) Reference collections (related to one project only)	Difficult to replace information (legal documents, email, project collections) Reference collections (related to several projects, web bookmarks) Other collections (passwords, address book)	Reference collections (a photo collection of a photographer, research data) Difficult to replace information (legal documents)
Personal incentive for preservation (little or no resources invested in creation)	Ephemeral information (to-do lists, browsing history)	Difficult to replace information (personal email) Other collections (personal address book)	Precious irreplaceable information (personal photos, videos, diaries, scapbooks, certain email) Other collections (personal address book)

The value of information assets from the economic perspective has extensively been studied in corporate environments. While it is easy to determine the value of preservation technologies, there is no easy way to answer what is the value of information in order to justify the cost of preserving it. In the corporate world information is often seen as knowledge rather than as a thing (objects that contain representation of knowledge) [7]. However, knowledge is most commonly preserved and managed only through information (as thing) [16]. Lavoie defined preservation as an economic activity and described three areas affecting its sustainability: distribution of responsibilities among relevant stakeholders, defining incentives (profit, kudos, decreased risk) for information preservation, and deciding on the organisation of preservation strategies to realise preservation objectives in the most efficient way [20]. Currall and McKinney pointed out that stakeholders need to understand three things: value of information, its technological fragility and the need for sustained support. In their model they divided value of a particular information item in four interrelated dimensions: (i) value information brings to costumers, (ii) financial value in terms of cost savings or income, (iii) value of information to the advancement of employees and (iv) organisation [9]. In their view the value is multifaceted and it needs to be re-evaluated over time (e.g. end of the legislative compliance) which is similar to a more general objective theory of value – valuing information based on observed and learned facts in the valuing context based on benefits it provides. While objective theory is more general, such corporate models provide a framework to assign specific metrics in each dimension based on the goals. This model can also be applied to personal information preservation goals [1] For example a family photo can bring advancement (pleasure, reminiscing, understanding family history) to its owner (employee), descendants (costumers) and when shared can tighten family connections (organisation).

Nonetheless, the motivation to preserve information can also be studied in the light of social science theories of values. One of such lists 10 basic interrelated values, varying in importance that serve as guiding principles in our lives: self-direction, stimulation, hedonism, achievement, power, security, conformity, tradition, benevolence and universalism [31].

Motivation for preservation and adoption of preservation technologies

Adoption of ICT has been accounted to both intrinsic and extrinsic motivations [11, 13]. With the former, the adoption of the technology is driven by self desires, interest and enjoyment. Whereas the extrinsic motivation has external factors in place (usually in the form of rewards or punishment) which incite the adoption (regardless if the intrinsic motivation is already present).

PIM activities are often viewed as a supporting activities linking information to other production activities. They have been described either as as a chance for a welcome break, or as a waste of time [5, p 175]. Users do not spend enough time on PIM and, as it has been shown in numerous studies, are mostly not satisfied with the state of their information spaces. Maintenance, as the way to improve this, gets least attention of all four PIM activities (see Section 2); there are usually no short-term outcomes associated. We have observed that backup activities (part of maintenance) are motivated (triggered) by external factors causing the “gut feeling” that something might happen to information (computers getting slower, before travelling, a lot of time and effort invested in information, etc.) or after information loss has already taken place. Backing up can clearly not be associated with intrinsic motivation. The “fear of losing” information drives users to backing up.

While needs and values are important in the motivation process of information preservation, gaining the sufficient knowledge and adopting steps in order to efficiently preserve information is crucial. It has been noted in many studies that people have a distrust for technologies which was also observed in this study. We already mentioned the importance of having a sense of control over software which is also one of the drives of intrinsic motivation (the flow of human-computer interaction)[13]. The perceived usefulness of a particular system (e.g. backup software) towards achieving better job performance is an example extrinsic motivation [10]. However, the question remains on how to motivate users to seek out these new technologies if personal information maintenance is not high on the priorities lists.

9 Conclusion

Being able to preserve information is as important as being able to create it. Even if preserving information in many cases takes less time than creating it, this study had showed that nearly a quarter of most valuable files and a third of most important folders at the time of answering the survey could not be (fully) restored after a disaster. The numbers are probably much higher for less important information as a great proportion of information that could be saved would be saved from unplanned backups. Manual, infrequent and selective backups triggered at the “gut feeling” moments or the lack of backups are the main cause behind possible information loss.

The presented survey sparked many reactions or as one of the respondents put it, this was *“A thought-provoking survey!”*. Several PIM studies reported to have triggered self-reflexive thoughts of PIM practices among participants (e.g. [6]). This study caused analogous response and many respondents commented that they had no idea how weathered their backup plans were. While the majority of

[1] Note that from a PIM perspective, information an individual creates and organises for work or private life is still personal – thus there are always two perspectives to corporate information from a corporate and from a personal point of view of its creator/curator.

respondents are conscious about the consequences of failing to do proper backups, the lack of motivation to improve current practices and the reliance on unplanned backups is impressive. Besides, respondents have an incomplete picture of implications behind using solutions they use and of possible available solutions.

While in the short-term the mess of infrequent (and manual) planned and unplanned backups is somehow welcomed it can hinder the long-term preservation possibilities. This mess of backup procedures is interleaved with dynamic, adaptive and fluid workflow of today's knowledge workers. It is thus a consequence of work routines, intersecting practices, and artefacts. To preserve important information and to fit in such work dynamics it is essential to develop backup solutions (hardware and software) that will be autonomous and give users' the control over its operation. One can say that technically the problem is solved as solutions are available to consumers. The question is why are we then still relying on casual stewardship?

References

[1] Randall J Allemang. Backup your data- the need for a new data archival format. *Sound & Vibration*, 45(1):5–6, 2011.

[2] Deborah Barreau. Context as a factor in personal information management systems. *Journal of the American Society for Information Science*, 46(5):327–339, 1995.

[3] Deborah Barreau and Bonnie Nardi. Finding and reminding: file organization from the desktop. *SIGCHI Bulletin*, 27(3):39–43, 1995.

[4] Francine Berman. Got data?: a guide to data preservation in the information age. *Communications of the ACM*, 51(12):50–56, December 2008.

[5] Richard Boardman. *Improving tool support for personal information management.* PhD thesis, Imperial college London, University of London, 2004.

[6] Richard Boardman and M. Angela Sasse. ”stuff goes into the computer and doesn't come out”: a cross-tool study of personal information management. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 583–590, New York, NY, USA, 2004. ACM.

[7] Michael K. Buckland. Information as thing. *JASIS*, 42(5):351–360, 1991.

[8] BUMI. Q2 2012 data backup: Industry survey results. `www.backupmyinfo.com/pdf/q2_2012_survey_results.pdf`, 2012. [Online; accessed 24-May-2013].

[9] James Curall and Peter McKinney. Investing in value: Investing in value: Investing in value: A perspective on digital preservation. *D-Lib Magazine*, 12(4):2, 2006.

[10] Fred D Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.

[11] Fred D Davis, Richard P Bagozzi, and Paul R Warshaw. Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of applied social psychology*, 22(14):1111–1132, 1992.

[12] Nicolas Ducheneaut and Victoria Bellotti. E-mail as habitat: an exploration of embedded personal information management. *Interactions Magazine*, 8(5):30–38, 2001.

[13] Jawaid A Ghani and Satish P Deshpande. Task characteristics and the experience of optimal flow in human—computer interaction. *The Journal of psychology*, 128(4):381–391, 1994.

[14] Sarah Henderson. *How do People Manage Their Documents? An Empirical Investigation Into Personal Document Management Practices Among Knowledge Workers.* PhD thesis, University of Auckland, 2009.

[15] William Jones. *Keeping Found Things Found: The Study and Practice of Personal Information Management.* Morgan Kaufman, Burlington, MA, 2008.

[16] William Jones. No knowledge but through information. *First Monday*, 15(9 - 6), September 2010.

[17] A.K. Karlson, G. Smith, and B. Lee. Which version is this?: improving the desktop experience within a copy-aware computing ecosystem. In *CHI '11: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 2669–2678. ACM, 2011.

[18] Joseph 'Jofish' Kaye, Janet Vertesi, Shari Avery, Allan Dafoe, Shay David, Lisa Onaga, Ivan Rosero, and Trevor Pinch. To have and to hold: exploring the personal archive. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 275–284, New York, NY, USA, 2006. ACM.

[19] Andy Klein. 10% now back up daily, 90% to go! `http://blog.backblaze.com/2012/06/12/10-now-back-up-daily-90-to-go/`, June 2012. [Online; accessed 24-May-2013].

[20] Brian F Lavoie. Of mice and memory: economically sustainable preservation for the twenty-first century. *Access in the future tense*, pages 45–54, 2004.

[21] Edwin A Locke. The motivation sequence, the motivation hub, and the motivation core. *Organizational behavior and human decision processes*, 50(2):288–299, 1991.

[22] Wendy E. Mackay. Diversity in the use of electronic mail: a preliminary inquiry. *ACM Transactions on Information Systems (TOIS)*, 6(4):380–397, 1988.

[23] Catherine C Marshall. How people manage personal information over a lifetime. In William Jones and Jaime Teevan, editors, *Personal Information Management*, pages 57–75. University of Washington Press, Seattle, WA, 2007.

[24] Catherine C. Marshall. Rethinking personal digital archiving, part 1: Four challenges from the field. *DLib Magazine*, 14(3/4), March 2008. Part 1 of a two-part article.

[25] Catherine C. Marshall. Rethinking personal digital archiving, part 2: Implications for services, applications, and institutions. *D-Lib Magazine*, 14(3/4), March 2008. ISSN 1082-9873.

[26] Catherine C Marshall, Sara Bly, and Francoise Brun-Cottan. The long term fate of our digital belongings: Toward a service model for personal archives. In *Archiving '06: Proceedings of Archiving 2006*, pages 151–156, 7003 Kilworth Lane, Springfield, VA 22151, May 2006. Society for Imaging Science and Technology.

[27] MSDN Microsoft. Windows 7 backup and restore deprecated. `http://msdn.microsoft.com/en-us/library/windows/desktop/hh848073.aspx`, November 2012. [Online; accessed 25-May-2013].

[28] Donald A. Norman. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, December 2003.

[29] Ayn Rand, Nathaniel Branden, Alan Greenspan, and Robert Hessen. *Capitalism: the unknown ideal*. Signet, 1966.

[30] Pamela Ravasio, Sissel Guttormsen Schr, and Helmut Krueger. In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 11(2):156–180, 2004.

[31] Shalom H Schwartz. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in experimental social psychology*, 25(1):1–65, 1992.

[32] Seagate. How much would you pay for peace of mind? `http://consumer.media.seagate.com/2012/08/the-digital-den/how-much-would-you-pay-for-peace-of-mind/`, August 2012. [Online; accessed 24-May-2013].

[33] S. Subashini and V. Kavitha. A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, 34(1):1 – 11, 2011.

[34] Symantec. Avoiding the hidden costs of the cloud. www.symantec.com/content/en/us/about/media/pdfs/b-state-of-cloud-global-results-2013.en-us.pdf, 2013. [Online; accessed 24-May-2013].

[35] Ludwig Von Mises and Bettina Bien Greaves. *Human action.* Liberty Fund, 1949.