Exploring the methods and practises of personal digital information archiving among the student population

Revised version of paper that was presented on BOBCATSSS conference 2015.

Hana Marčetić

Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences, Department of Information Sciences, Osijek, Croatia

Recenzenti:
Mgr. Eliška Pavlásková
Prof. PhDr. Jela Steinerová, PhD.

Abstrakt:
The issue of personal archiving is one often raised in the context of the fast changing digital era. Even among the "informationally literate" personal digital archiving is more often simply a side-effect of generating content in the digital environment than a planned activity. The aim of this paper is to research the extent to which the student population employs the doctrines of digital curation, digital preservation and digital stewardship. This group is interesting for examining in this context because of the unique mixture of formal, mainly scholarly, and personal information that they govern in their day to day activities and in digital formats. Some of the information that this research attempted to obtain concerned the actions individuals undertake to acquire, store and conserve digital objects, the formats they use and the practices they employ in the process. An attempt was made to determine whether the migration of content is practiced, where and how digital information is stored, as well as whether or not scholarly objects (such as those used for study) are handled differently than informal ones (such as those used for entertainment and other private purposes). The paper also provides insight into student practices regarding the creation of digital copies of objects and the use of other media such as portable memory sticks, CDs, DVDs and Blue-Ray discs. The research touches on a number of other questions, such as: how often do students revisit the contents once they've created them? Do they use digital archiving in the cloud? Do they create backups for the objects they have identified as important to preserve? This paper also contributes to identifying the specific criteria for the selection of digital information that will remain preserved for the future other than pure chance. The survey was conducted on undergraduate and graduate information science students from the Universities of Osijek, Zagreb and Zadar. Online questionnaire sent to said institutions and dispatched among information science students resulted in data that can be used in attempts to answer the above mentioned questions. Once the survey was closed, the gathered data was analysed using SPSS programme and used to draw conclusions regarding personal digital archiving practices and awareness among this population, along with other interesting findings this study brought up.

The results of this study imply that information science students in Croatia are quite aware of personal digital information issues in general. However, when it comes to managing collections constructed through accumulation of personal digital information, certain practices such as migration, deleting documents no longer needed, and so on, seem to be omitted, at least to some extent. Organising practices and formats most often encountered imply that this population is subject to the same challenges personal digital archiving brings to all users technology, such as unclear selection criteria and inconsistencies in methods practiced. That being said, it is evident that the level of awareness improves the attitude towards the PDA issue and increases the chance for this issue to be dealt with suitably. It should also be noted that the phenomena of information fatigue is widespread among the respondents regardless their differences in general.

Keywords: *personal archiving, digital documents, digital preservation, personal information management*

Introduction

Personal archiving is an important part of the concept of personal information management, which implies acquiring, organizing, using and preserving information for personal use. It is then no wonder that, since entering a new era characterized by a great amount of content generated on a daily basis, personal digital information (PDI) archiving has emerged as a topic vividly discussed. The intention of this paper is to provide a preliminary insight into the extent to which the students of information science differ in their PDI archiving habits, compared to other examinees of previous studies. There have been studies intent on examining students' habits and actions regarding the information they interact with in different circumstances, but none on a sample as is chosen in this study. A scarcity has been detected in the existing researches of individual practices of digital archiving among students of information sciences. This is the population that is presumably going to take on dealing with the difficult and complex issue of digital curation in whole, as well as personal digital information archiving, its vital part. Not only is this a pressing issue as is, but it is also safe to presume that it will only increase in complexity with time. As future employees of memory institutions, workers in information technology sector and educators in the field of digital information managing, information science students' skills and practices in this area should be of interest to the knowledge society. It should strive to educate their future information experts in such a way that they be able to best tackle the ever increasing issue of digital archiving. Young adults currently being educated in the field of information science are the ones who are in the best position to acquire the skills of dealing with the rising tide of data and overwhelming amounts of information with most success. No other group of people is in a better position to take on dealing with this issue. Not only should the society in whole take interest in the level and the quality of education they acquire during their studies, but attempts should also be made to research their skills and practices in personal digital archiving, so as to determine their current attitudes and their level of familiarity with related issues. In the words of Sarah Kim "we are heading toward an increasingly individually-focused culture where each individual becomes a central unit of social action" (Kim, 2010, p. 47). This is why this paper argues it imperative to convey the importance of personal digital archiving within this population, which carries upon it the mark of the digital era. It is hoped that it will contribute to raising awareness about this question.

Literature review

Regardless of the level of awareness, the urgency of preserving digital personal information is rarely recognized by individuals enough to induce a concrete and systematic care for it. Marshall (2007) claims this is a result of a still prevailing technological optimism, that is, the belief that the problems will resolve themselves in time, and with progress. However, to anyone remotely interested in the problem of PDI it is clear that the only way of preserving digital information is actively engaging in its preservation. Many studies and papers tried to tackle the problem of what those actions should be and how individuals should approach them. The process of PDI archiving is set in motion as soon as an individual has encountered information and decided to keep it for future use. As Bruce, Jones and Dumais (2004) assert, „Keeping acts are interventions by the individual". However, often „a leaving method will be preferred under the assumption that a person in question will be able to locate the information source again if a need arises", the same authors point out. Williams, Rowland & Leighton John (2009) noted a large number of keeping methods, such as sending an e-mail with a URL to oneself or others, saving the entire Web page or printing it out, bookmarking it, or pasting a link into a separate document. Bookmarks and favourites were identified as the most common methods of keeping information. The authors also observed that often people will rely on re-accessing information by searching again for the desired Web information, entering URL from memory (using browser suggestions as help) and reaching it from a known point of access (i.e. a Web portal). This approach is known as leaving methods. Certain questions from that study revealed that most examinees overlooked the steps of transferring or deleting files they no longer needed despite being aware of the future complications it could induce in time (Williams, P.,

Rowlands, I., Dean, K. & Leighton John, J., 2008). The authors go on to conclude that the perceived value of a digital artefact will determine whether an individual will attempt to preserve it by creating a backup version and/or creating a hard copy of it. In a previously mentioned paper by Marshall (2007) the author detects that, in addition to skipping the step of creating backups of their documents and digital artefacts in general, people are also quite unfamiliar with formats available for saving them. She further hypothesizes that the tools for PDI archiving that are growing in number and availability in the current digital environment are the best strategy currently available for providing the favourable circumstances to successfully archive digital artefacts. These findings are compared to data collected from Croatian students of information science who participated in this study.

For students in general, PDI archiving intertwining the personal and formal spheres of digital lives presents a complex problem. They are often unaware of the amount of content they interact with and generate on a daily basis. This situation is not promising for the future accessibility and usability of different artefacts that could be of interest, to both the creator of the content and other potential interest groups. When one considers the PDI environment youth functions in, a problem of increasingly fragmented methods and applications used for archiving arises (Robinson, S. & Johnson, F., 2012, p.3**).** This fragmentation is visible not only in the multiple technologies and versatile gadgets accessible for generating and accessing different content, which in turn provide a wide variety of platforms for archiving options, but also in different strategies, multiple social networks, tools and cloud systems used, along with many other options an individual chooses from with each attempt at PDI archiving. The information is scattered over different platforms and units such as CDs and USB memory sticks, social media sites such as Facebook and Flickr and in many variants, various file formats and so on (Korhonen, M., 2013. p.85). This is why Beagrie (2006, p. 12-13) argues that the only way for society and individuals to preserve what he calls digital knowledge useful and useable is to continuously update, maintain and access it, as well as put significant effort into long-term preservation.

The enormous amount of information has earned today's youth the nickname of Generation C (for content). And the extent of it continues to grow. A case study examining how graduate students manage digital scholarly articles found differences in practice over different subject disciplines and between genders, and also the aforementioned use of multiple approaches and management practices (Huvila, I., Eriksen, J., Häusner, E. & Jansson, I., 2014). All of the mentioned findings have been taken into account in the attempt to determine existing distinctions in practices of PDI archiving by information science students in Croatia.

Personal digital information archiving among the information science student population

Aim, purpose and methods of the study

This paper is based on the results of a survey conducted on a sample of 220 information science students from three different universities in Croatia: the University of Zagreb (the Department of information and communication sciences), the University of Zadar (the Department of information sciences) and the University of J. J. Strossmayer of Osijek (the Department of information sciences). This approach was chosen in an attempt to explore the practices and contributing factors of personal digital archiving among future information science experts, their level of awareness of this issue and the way their education influences their practices regarding the matter. Their practices and opinions were explored via an online questionnaire, in order to answer the following research questions: Are future information experts aware of

the problems of personal digital archiving? What actions do they engage in trying to preserve their digital artefacts? And finally, do they feel the amount of content weighing down on them?

Based on the previous research discussed earlier, in this study it is hypothesised that:
1. the level of education and awareness of the importance of information for today's world increases and
2. gender a) influences the level of awareness of the importance of personal digital archiving,b) induces more effort to be put into organizing personal digital artefacts and c) helps lower the feeling of information fatigue.

This paper aims to raise the questions of the imperatives in personal digital archiving. It is hoped it will bring about further research that will give better insight into the problem of this specific interest group, for who should be aware and prompt the explicit and tangible actions in this field if not the young information scientists belonging to „Generation C". The objectives of the conducted study concern the attitude these groups have towards the content they generate and how they endeavour to preserve their digital artefacts specifically.

Results and discussion

General information about respondents

Considering the sample chosen for this research it is not surprising that the number of female participants exceeds the number of male ones; 79,5% (175 participants) were female, while 20,5% (or 45 participants) were male. Most of them, 63,2% (139 participants) claim to have taken classes regarding information and data preservation in the course of their education. Data regarding their level of education and programme were also gathered (Table 1). The results were then compared to the questions concerning the level of awareness with respect to the care over personal digital artefacts, the effort put into it, and subsequently, the level of information fatigue felt; 47,3% (N=104) feel it, 20,5% (N=45) don't and 32,2% (N=71) don't think about it.

Table 1 – Participants by department affiliation

Department	Undergraduate level	Graduate level
Department of information sciences, Osijek	28,2% (N=62)	11,3% (N=25)
Department of information sciences, Zadar	6,8% (N=15)	13,2% (N=29)
Department of information and communication sciences, Zagreb	15,5% (N=34)	25,0% (N=55)

When asked whether they think taking care of digital documents in their personal collections was important, a great majority answered affirmatively, namely 94,5% (N=208) while 5% had no opinion on that subject (N=11). Only one person considered it unimportant, i.e., 0,5%. A similar situation occurred regarding the question of the effort put into organising, when a vast majority of 90,9% (N=200) again agreed that they strive to organise their digital belongings, while only 2.3% (N=5) of the sample claimed not to put any effort into it. The remaining 6,8% (N=5) neither confirmed nor denied any effort. This implies that the students of information sciences in Croatia are generally conscious of the impact their digital lives have on their digital collections and at least make attempts to keep them organised.

However, when it comes to managing their collections further down the line, it seems they fall behind a little. For example, only 44,5% (N=98) claim to delete documents they no longer need regularly, but the numbers go up 47,7% (N=105) when those who claim to do it „sometimes" are taken into account. Again, a much smaller percentage of 7,7% (N=17) claim not to ever take

this action. This may, along with the fact that this is a hectic era of immeasurable amounts of digital contents, be part of the reason why number change gravely when it comes to information fatigue.

Migration and data management

Revisiting previously saved data and checking it regularly is one of the actions most often mentioned when it comes to specific activities attainable for prolonging the lives of artefacts constituting digital collections (Table 2). However, being very time-consuming, this crucial practice is often either completely overlooked or done insufficiently enough, without premeditation.

Further on, migrating materials from medias no longer used is equally unorganised and lacking a strategic approach (Table 3). Even among those who recognise its importance and practice it, it's apparent they only do it with materials they themselves find important. This, in turn, brings up the question of selection criteria. Also, who can claim with certainty that some information will not be considered important in the future?

Table 2 – Digital data and collection management

I regularly check all the data once saved	12,7%(N=28)
I regularly check the data once saved only if I find it important	32,3%(N=71)
I only check saved data when I need it.	53,6%(N=118)
Other	1,4%(N=3)

Table 3 – The practices of migrating digital data

I strive to migrate the content from old media	16,8%(N=37)
I only practice data migration with the most important information	54,5%(N=120)
I migrate information randomly, without any criteria	16,8%(N=37)
No, I don't consider it important	9,5%(N=21)
Other	2,3%(N=5)

Only one (0,5%) of the participants answered that she doesn't practice migration, as she said, „although I think I should". Others have claimed they can find anything they need online or that they probably already have it somewhere on their computers, so they don't see why they should migrate from old medias.

Another, perhaps more familiar method in digital information archiving is creating backup. Again, backup will mostly be created for information individuals estimate consider valuable enough to put in the effort (official documents), or they will just randomly choose „on the go" without applying any particular strategy (Table 4).

Table 4 - Backup

I backup all my content	15,5% (N=34)
I only backup official documents	26,8%(N=59)
I only backup my personal photos and videos	7,7%(N=17)
I create backup sometimes, randomly	25%(N=55)
No, I have no need for it	25%(N=55)

Organising practices

In order to gain insight into the ways information science students organise their collections, the participants were asked a series of multiple choice questions regarding the actions they take to organize documents, as shown in the Table 5 below:

Table 5 – General organising practices

Organising in folders	94,5% (N=208)
Adding metadata	36,4%(N=80)
By separating formal from informal data	47,3%(N=104)
Using a tool	0,5%(N=1)
Other	N=0

Purportedly, most of the participants organised their documents in folders. Detailed actions are represented in the Table 5.1 below, expanded with comments implying alphabetic, enumerative, theme and date-related strategies. Only one person claimed to have used an indexing tool in document organisation, called „Where is it? ".

Table 5.1 - Organising in folders

Organising method	**Percentage (frequency)**
By date	22,7% (N=50)
By document type	54,5% (N=120)
By importance	77,3% (N=170)
Other	2,7% (N=6)

Table 6 represents a comparison of actions practiced depending on the nature of the content. Others have answered that they don't download content from the Internet at all, but simply access it online when the need arises.

Table 6 – Digital data management: comparison

Digital data management actions	**What do you do with files you think you might need again**?	**Where do you file copyrighted content you downloaded from the Internet (music, films, books, etc.)?**
I file them on my PC.	15,5%(N=34)	89,1%(N=196)
I make several copies and save them to different devices (e.g. an external hard-drive, another computer...)	35%(N=77)	16,4%(N=36)
I transfer them to a cloud (e.g. Dropbox	31,8%(N=70)	5,9%(N=13)
I bookmark or pin it in my browser.	39,5%(N=87)	16,8%(N=37)
I email myself an attachment or the URL.	50%(N=110)	10%(N=22)
I save the URL to a file on my PC.	15,5%(N=34)	8,2%(N=18)
I print the hard copy.	20,9%(N=46)	N=0
I save them to another medium (CD, DVD, Blue-ray)	15,9%(N=35)	19,5%(N=43)
I copy the most important information to a separate file	19,5%(N=43)	N=0
I save the entire Website to my PC	9,1%(N=20)	N=0
Other	0,5% (N=1)	1,5% (N=3)

Social platforms

Table 7 represents answers to the question of which social platforms participants use for the storage and organisation of photos. Other suggested platforms included Tinypic, Ptičica, Picassa, Imgur, DeviantART and Fotozine.

25,5% (N=56) of participants claim to use social platforms in data organisation, 25,9% (N=57) only use it when informal information is in question, while as much as 48,6% (N=107). Organising practices in this environment are quite different from those previously discussed. Different options are available and utilised, as shown below. Other answers included „I don't organise pictures", „I send them to someone" and „I tag it with my surname (e.g. #surname)".

Table 7 – Social platforms used in organising photos

Flickr	2,3% (N=5)
Tumblr	8,6%(N=19)
Facebook	59,1%(N=130)
Twitter	2,7% (N=6)
Instagram	14,1% (N=31)
Pinterest	5,5% (N=12)
None of the above	28,6% (N=63)
Other	5% (N=11)

Table 8 – Organising in social platforms

Organising in folders	50,5% (N=111)
By tagging people	38,2% (N=84)
Using hashtags (#)	10% (N=22)
Adding descriptions and annotations	24,5% (N=54)
Other	10,9% (N=24)

Formats

44,1% (N=97) of participants pay attention to formats used in data storage while 55,9%(N=123) don't. For text documents, the most frequently used format is PDF, photos are mostly saved as JPG documents, while MP3 and AVI are most frequently used for long-term preservation of audio and video content. Detailed answers are presented in tables 9 to 12.

Table 9 - Textual documents

Format used	Percentage (frequency)
PDF	40,5% (N=89)
DOC	35% (N=77)
ODF	0,5% (N=1)
TXT	13,6% (N=30)
HTML	3,6% (N=8)
RTF	2,3% (N=5)
Other	0,5% (N=1)

Table 10 – Photos and visual content

Format used	Percentage (frequency)
JPG	40% (N=88)
JPEG	27,7% (N=61)
PNG	20,9% (N=46)
BMP	5% (N=11)
TIFF	2,3% (N=5)
GIF	0,5% (N=1)
Other[1]	2,9% (N=6)

[1]Participants also suggested CRW, PSD, Photoshop RAW and WEBM fomats.

Table 11 - Audio content

Format used	Percentage (frequency)
MP3	41,8% (N=92)
MP4	19,5% (N=43)
WMA	14,1% (N=31)
WAV	6,8% (N=15)
AIF	0,5% (N=1)
AU	N=0
Other[2]	2,3% (N=5)

Table 12 - Video content

Format used	Percentage (frequency)
AVI	34,1% (N=75)
MPEG	33% (N=15)
DVD	10,5% (N=23)
DIVX	9,5% (N=21)
WMV	6,8% (N=15)
MKV	13,2% (N=29)
Other[3]	1,8% (N=4)

Conclusion

To some level, consciously or not, all the participants in this study are compelled to archive their personal digital belongings in the form of different information and objects. However, the only way of preserving digital information is actively engaging in its preservation. While it is evident that information science students are widely aware of the issue at hand and tend to consciously attempt to manage and organise their digital lives, some aspects (such as migration or using tools for PDA) seem often to be overlooked. The gathered data show no correlation that the effort put into organising digital information is in any correlation with the university attended, but it does seem that graduate students are more prone to attempts of organising it than undergraduate students and, interestingly, so are female compared to male participants. Students who have already taken classes affiliated to the subject of personal archiving and preservation are not only more likely to try to organise their materials, but also agree to a greater degree that this is an important issue. The year they're currently attending, the university they're enrolled in, and gender differences do not, however, seem to influence the level of awareness on this matter significantly. Data also implies that everybody, regardless of factors, feels the same amount of information fatigue.

The question of the level of awareness regarding PDA issues, amount of effort put into PDA practices and the feeling of the so called information fatigue were additionally analysed in SPSS program for statistical analysis. Cross tabulation analysis and chi square tests were carried out to compare above mentioned questions to gender, level of education, attended university and previously taken classes. After such analysis following conclusions can be inferred with regards to hypotheses postulated earlier in the paper:

- the number of classes taken during the education and the level of education (graduate or undergraduate) seems to influence
  - a) the level of awareness of the importance of personal digital archiving,
  - b) the effort put into organizing personal digital artefacts, and
  - c) helps lower the feeling of information fatigue,
- while gender
  - a) doesn't seem to have any influence on the level of awareness,
  - b)appears to increase the amount of effort, put into organising personal digital information (female participants put in more effort), but
  - c) doesn't have any apparent influence on the feeling of information fatigue.

[2] Participants aslo suggested FLAC format.

[3] Participants also suggested MP4 format.

Several questions arise for further research. Perhaps the most evident one is the question of selection and evaluation process of deciding which digital possessions to actively manage and engage in its preservation. It has been noted that a significant number of participants repeatedly mention importance as the main criteria for deciding whether or not a certain action will be undertaken in order to ensure the longevity of a particular digital object or piece of information. It would therefore be interesting to further look into the process of deciding whether to label a certain item important or not. Another problem poses the fragmentation of methods used for PDA among students, along with the understandable notion that the more effort a certain action requests, the lesser the likelihood it will be carried out. Also, a conclusion could potentially be drawn from the migration practices data gathered during this research combined with those regarding the format choices, that much of the longevity from the technical viewpoint is left to chance. In other words, the combination of popularity of reliable formats and durability of the form in which an object is initially saved determines its odds for preservation. In the literature review it is mentioned that a case study by Huvila et al. observed differences in PDA in relation to the gender of participants, but the only difference observed in this sample referred to the level of effort put into PDA. A majority of the participants concluded that social networks are not reliable for storing the information once it is evaluated as important. In addition to that, hard copies are still considered the most efficient way of storing objects when it is imperative to assure longevity. Moreover, from the psychological viewpoint this issue holds a great appeal for discussion. It would be interesting to further systematise the sense of information fatigue combined with character traits influencing the PDA practices and incentives.

This is only an outline of the personal archiving habits of Croatian information science students. It is recommended that a more extensive qualitative study be conducted on this subject.

References

Beagrie, N. (2006). Digital curation for science, digital libraries and individuals. *The international Journal of Digital Curation*, 1(1). DOI: http://50.17.193.184/omeka/files/original/800dc0599686ce2f20cc0e1c825d2889.pdf

Bruce, H., Jones, W. & Dumais, S. (2004). Information behaviour that keeps found things found. *Information Research*, 10(1). Retrieved from http://InformationR.net/ir/10-1/paper207.html

Huvila, I., Eriksen, J., Häusner, E. & Jansson, I. (2014). Continuum thinking and the contexts of personal information management Information Research, 19(1) paper 604. Retrieved from http://InformationR.net/ir/19-1/paper604.html

Kim, S. (2010). Personal digital archives: Preservation of documents, preservation of self. Retrieved from http://srhkim.com/files/PhD-proposal-final-sarahkim-dissemination.pdf

Korhonen, M. (2013). Private digital archives – Lost cultural heritage? In Niemi, H., Salo, S., Sipilä, S., Gustafsson, D. Linnovaara, K. (Eds.), *Essays on Libraries, Cultural Heritage and Freedom of Information*. Retrieved from http://www.academia.edu/5575711/Private_Digital_Archives_-_Lost_Cultural_Heritage

Marshal, C. C. (2007). How people manage personal information over a lifetime. In W. P. Jones & J. Teevan (Eds.), *Personal Information Management.*Retrieved from http://www.csdl.tamu.edu/~marshall/PIM%20Chapter-Marshall.pdf

Robinson, S. & Johnson, F. (2012). The process and affective environment of students' personal information management. *Enhancing Learing in the Social Sciences, 4*(2), DOI: 10.11120/elss.2012.04020005

Williams, P., Leighton John,J. & Rowland, J. I, (2009). The personal curation of digital objects: A lifecycle approach. *Aslib Proceedings, 61*(4), 340-363. DOI: http://dx.doi.org/10.1108/00012530910973767

Williams, P., Rowlands, I., Dean, K. & Leighton John, J., (2008). Digital Lives: Report of Interviews with the Creators of Personal Digital Collections. *Ariadne 55*. Retrieved from: http://www.ariadne.ac.uk/issue55/williams-et-al