

Preservation is no Preservation without Proper Education

Anssi Jääskeläinen; Mirja Loponen; Mikkeli University of Applied Sciences; Mikkeli, Finland

Abstract

Public administration has a very strict obligation to preserve documents according to many laws and regulations. However, associations, private- and business sector does not belong under this obligation unless they are performing statutory mission. According to accounting obligation, every quarter that performs business or professional activity is accountable. Still in a worst case scenario, the only preserved document might be the mandatory one, which is stored on active hard disk. Archiving technology as well as information technology advances rapidly, but IT devices suitable for long term preservation do not belong to the off-the-shelves category. Lots of micro history (information, data, material and physical objects) disappear due to lack of people who possess enough know-how to understand the relevance and importance of preservation for the future. For example, 20 years of information from the Karelian evacuees disappeared from one local association after the “archivist” passed away and heirs get rid of all the “junk” that was found at the attic. The final obstacle to the preservation, in spite the technology, might be lack of a concrete plan and way of storing something. This again culminates in lack of a properly educated person who could establish an applicable preservation plan and obtain the required devices or services for required actions. From these starting points we at the Mikkeli University of Applied Science (MUAS) are aiming towards general preservation education that would greatly increase the preservation foundations of citizens. This paper and work behind it was conducted in an EU project funded by ERDF (European Regional Development Fund) programme for Eastern Finland

Introduction

It should be unnecessary to start introducing digital preservation to this audience. However the relationships of Finnish public administration, private and business sector actors and education to preservation must be introduced thus these forms the starting points of this paper.

In many sectors, the influence of Copyright Act and the Archives Act, in the Finnish legislation is strict. For example all following record creators: “Government offices, institutions, courts of law and other organs applying the law and other government authorities, municipal authorities and organs, The Bank of Finland, University of Helsinki, Social Insurance Institute and other independent public institutions, Government and municipal enterprises, The Greek Orthodox Church of Finland and its congregations and other collective bodies, organs and persons carrying out public duties according to acts and decrees” [1] must abide The Archives Act. In addition this strict legislation also reaches associations that conduct statutory mission; an example of such an association is Mannerheim League for Child Welfare. Parties that are obligated to preserve documents tend to have enough competence and knowhow to handle the act of preservation

whether it is digital or old fashion paper. However, according to authors’ opinion hands of these parties are too bound laws and rules created by actors who still live in the world of paper. Naturally certain level of bounding is necessary to keep all mandatory documents in safe, but too much laws, rules, regulations exceptions are a perfect way for killing inventions and evolution. For example, if The Ministry of Finance sends a request for comment to every municipality in Finland, they all record a new case, which is preserved with the request for comment. It makes no sense to preserve the same document for 336 times. Instead, it causes substantial costs and is very inefficient. However, law says it has to be done. Transition from document management to information management, where one information process would only have one owner with record and preservation responsibility, would make the process much more efficient and cost effective. Therefore, our suggestion is that the education designed to be given at second (and lower) level should focus on giving the picture of information management and preservation as a whole and not to consider some individual aspect too thoroughly. Gained knowledge could then be deepened at the university level.

In spite the list of ‘strictly bound by law actors’ is large, it completely rules out private and business sector as well as individual people. These parties that are not so strictly confined are the main target audience that requires proper knowledge derived from proper education. Hypothetical estimation by authors is: If one asks one hundred Finnish SMEs (Small and Medium Enterprises) to tell how they preserve their data, 95 % or more of them would say something like “We have this USB drive”, “RAID (redundant array of independent disks)”, “We store our data to DVDs”, “Our hard drives have three years warranty”, “We use Google Drive or SkyDrive”, “We use network drive”, “We have network file server”, “I don’t know”, “Why should we store old data?”, “What is preservation?”. From the given hypothetical answers, few could be considered as a form of preservation and in fact the authors might even recommend e.g. clouds, network drives, network file servers or raid for SME:s due to their benefits and simplicity. However, most of the answers clearly indicate the lack of knowledge and knowhow of preservation and its long term benefits. E.g. most people don’t know that the warranty of the hard disk will not cover stuff that has been stored on the hard disk; it only covers the physical device.

In order to start considering the development of education, some form of knowledge about the educational structure of the target country is required. In this paper, the Finnish educational system is briefly introduced to be able to understand suggestions and considerations thrown in this paper. Another important aspect when the education is considered is the age when the production of digital content starts. With modern technology, this age has become down quite a bit.

The ultimate intention of this long journey behind this paper is to provide knowledge and methods for digital preservation that

are in balance with the actual needs of the users whether they are individual people, SMEs or even governmental actors.

Finnish educational structure

According to the Basic Education Act "Compulsory schooling shall start in the year during which the child turns seven. Compulsory schooling shall end when the basic education syllabus has been completed or ten years after the beginning of compulsory schooling" [2]. After the mandatory part, pupils can continue with different paths, which include but are not limited to high school, vocational school or work life. Later on it is possible to continue studies at university of applied sciences or universities. Figure 1 describes this structure thoroughly. The same figure also presents the amount of students at 2011 inside the each education level according to the statistics Finland. It needs to be mentioned that this figure still uses old polytechnics terms, but the correct term is university of applied sciences.

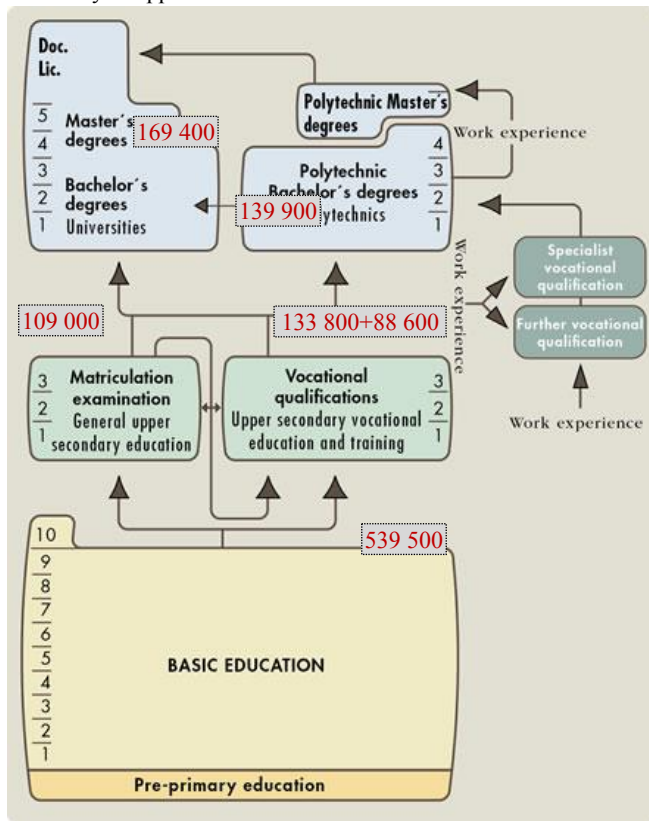


Figure 1: Finnish educational structure [2]

The Finnish national board of education decided the syllabus for the basic education in 16 January 2004. This level of education is mandatory for every citizen and therefore it is a relatively time consuming process to change the syllabus. It is not in the scope of this paper to go into this process, but it might not be the best alternative to start pushing preservation education into the national basic syllabus. It however, might be plausible to introduce it as an alternative course for those who are willing to take it.

The education in the field of archiving and archives management in Finland is currently under revolution, since the

NAS (the National Archives of Finland) announced that they won't continue their higher archive education after the year 2014. This education has already been partially transferred in to universities, but after 2014 it remains to be seen which one will welcome this responsibility. What can be guaranteed is the fact that MUAS will continue to give its higher level education in the field of electronic services and digital archiving since our NAS independent degree level program is approved by the Ministry of Education and Culture. MUAS degree level study belongs to the upper most part of Figure 1; however the intake is only about 20 students in every two years.

Therefore, the majority of the Finnish students won't ever even hear the terms 'preservation' or 'archive'. This means that the amount of graduated students who have adequate knowledge form the field of preservation is ridiculously low. Those lucky ones who have chosen to encounter it won't meet it until they are at least 18 years old. As a backup solution, some organizations and associations give archive and preservation trainings that are subject to noticeable charges.

Digital content

The described situation is very problematic thus according to the EU kids online study [3], in 2010, 96% of the Finnish 9-16 year old children and youths use internet at home, from which 77% on daily basis. At the European scale, the numbers are somewhat lower. Being in the internet equals producing content such as e-mails, social media writings, uploaded photos, chat, etc. Obviously most of the produced material is not worth preservation, but naturally there can be exceptions. 15-year-old Neil Ibatata found out that dwarf satellite galaxies share the same dynamical orbital properties and direction of angular momentum [5], For example. However, lots of valuable micro history produced by children, youths and other people who are not familiar with preservation, disappears.

Obviously, it is crucial to reach the basic knowhow of handling and preserving digital content in to the younger section. Furthermore, challenging the content of the internet is an important aspect in understanding what might be worth preservation. The best alternative would be to start this education (or could it be called customs training) as early as preschool age since in many comprehensive schools the digital devices, such as computers, tablets, and mobile phones are in everyday teaching use. However, as long as the compulsory basic education is out of our reach, the authors' suggest the second best alternative. Preservation education should first be introduced in upper secondary schools as part of either general social studies or as part of optional information technology courses. In vocational education schools, the preservation education could be given as part of a suitable training program. This education has already been designed in a national working group and an exploratory study of embedding this education into teaching is currently under way in co-operation at MUAS and South Savo Vocational College. With the aid of this education, about half of the Finnish youths would get the basic knowledge of digital content, backup solutions, preservation and some sort of impression of what might be worth preservation a from larger point of view.

Backup or preservation

According to Gibson [6] - data that is posted on the internet should be regarded as permanent after 20 minutes, even if the cases in which the originator has deleted the file. Nevertheless, does this mean that the data is secured, backed up, archived or something different? Gibson [6] also states that, the FBI now considers internet-posted data as permanent. They are relying to the fact that the data is cached by search engines for instance and once cached it is very difficult or even impossible to delete completely.

The other primary author of this paper has an extensive knowledge in the field of telecommunication and can confirm that the data on the internet is virtually permanent. However, it requires a lot more than average internet user capabilities and tools to dig up the required long-lost information. Backup on the other hand equals data protection and recovery. You could for example take a backup image of your hard disk and in case of accidental format or virus infection just recover the hard disk content from the image. An archive provides data retention and retrieval, as well as proof of chain-of-custody against every record, file, etc. Furthermore, a digital archive places many requirements that a simple backup cannot fulfill. Classification and categorization of information as

well as the essential component of every DRM (Digital Records Management) system, the destruction of information are missing from plain backup solutions. Backup solutions won't take any actions to ensure long-term readability and accessibility of the stored information about while digital archive handles these automatically. Finally, secured data on the other hand means the data is accessible only to those people who have a permission to access it.

The data on the internet is permanent, but it is not backed up, archived or secured, it is just cached. Cache might be a good utility when cyber-crimes are being solved, or some not properly backed up or archived data is accidentally deleted, but it is not a way to backup or preserve anything since the author of the information does not have any influence on how cache works and what it does.

Table 1 below identifies the cons and pros of different storage approaches. The same things that could be included in the proposed second level training. While the table is extensive, it still reveals only part of the truth, with optical media, USB sticks, external drives, network drives, network file servers and raid it is mandatory to replace the utilized storage e.g. after the warranty period is over or when it brakes. This naturally increases costs.

Table 1: Cons and Pros of utilized "storage" solutions

Solution	Pros	Cons	Cost Tb (€)	Utilization
Nothing	*No extra costs	*One disk failure and everything might be gone	0	If you don't mind about losing your data
Optical media (local)	*Common devices *Common media *Simple to use	*Media life *Scratches *Limited space	*80-1500€ (dvd-r) *160-400€ (blu-ray)	Transferring data, storing software installation packages, movies, music
usb stick (local)	*universal connection *cheap *utilize as hard disk *lifetime (if physically ok)	*vulnerable to physical impacts *cheap imports	500-1000€	Transferring data, temporary storage, presentations, currently one of the most used ones among end users
external usb disk (local)	*universal connection *utilize as hard disk *Cheapness / per Tb	*Life time *Should be mirrored in case of failures	40€→	Among usb sticks, one of the most used ones. Install OS again but safe all other data.
Raid (local)	*Totally transparent for user *Speed benefits on certain raid levels	*Set up might require IT professional *Requires disk buffer *If raid controller brakes, might require new from same manufacturer	Raid 1, 80€→ Raid 6, 160€→	Most used among individual IT-nerds and small enterprises.
Own digital archive (local)	*All in own hands	*Costs of operation	a lot	Option only if it is also an intention to sell as service
Network drive (semi local)	*Utilize as hard disk *Not dependable of PC hard disk failure	*Should be mirrored or regularly backed up	50€	Home network, backup, media streaming
Network fileserver (semi local)	*Fully independent unit, *Serves also in other purposes	*Should be raided or backed up to ensure data protection	2500€ →	Office network
Digital archive (Remote)	*All benefits of true archive	*Costs	Depends on many things (100-450€/month)	Data must be safe no matter what, privacy & security related content
Cloud (Remote)	*Utilize as hard disk *Access anywhere *Data is secured and accessible *Supports metadata	*Physical location of file is unknown *Most likely foreign access and usage rights contracts	40€/month → but 2-25 gb free	Various different purposes, transferring data, storing, backup, personal archive.
Social media (Remote)	*Access anywhere *Share with friends	*You have probably given all rights to the service provider	free	Youths

Cloud storage services, in spite working on the internet are different. These provide transparent long term data preservation, security and compliance for basic end users. While the legislation of certain countries prohibits the utilization of these clouds for governmental purposes for instance, the advantages of clouds for organizations and individuals are evident when compared with the true digital archives or offline backups. The cost of storage space and monthly fees are smaller than with digital archives and not particularly larger than with offline USB drives, for instance. Therefore, instead of trying to convince people and organizations to use true digital archives they should be convinced to resolve what might be the best option for them. This knowhow and will are currently missing and could be provided with a proper education.

In authors' opinion USB hard drives or network drives are a noteworthy solution for individual end users and maybe even for small companies where the amount of produced information is relatively small. However, when the amount of crucial or business critical data is larger, minimum amount of security should be created e.g. with clouds, suitable RAID or network file server. Public cloud services are easy to utilized, since they are device independent, commonly equipped with sophisticated highly developed UI and are simple to take into use. However basic cloud drive agreement commonly states something like 'user may not use the service to store, transfer or distribute content on behalf of third parties'. Nevertheless, there probably can be some organizational agreements that permit more uncontrolled usage. Another issue with the public cloud services is trustworthiness, but if this worries, it is always a possibility to set up your own private cloud.

Unfortunately, human mind does not wish to consider the worst case scenario and in the best case scenario, backed up or stored information is newer even needed. Therefore it could be incredibly difficult to convince some CEO (Chief Executive Office) of SMEs why they should purchase 3000€ system or pay certain amount per month instead of buying a bunch of cheap external USB hard drives since those have been working just fine so far. As an example of not considering the preservation of old material, a representative of Lego Inc. told in a seminar that Lego had a really hard time in winning a lawsuit against a Chinese manufacturer back in the days when they did not maintain their own archive of used packages and molds. After that incident, The Lego Idea House was founded and it stores a model of every used package, brick mold, etc. In addition to confidence it brings, it also serves as historical exhibition that has brought more business partners and clients for Lego.

This does not mean that the authors are suggesting that every party should build and maintain their own archive. Instead we suggest that you should all take at least some time to resolve your current data or other possessions and create e.g. a simple SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis to see whether you should take some actions or precautions maybe in a form of education.

Finally if the preserved material is truly valuable, privacy sensitive, considers medical records, confidential and it is likely that someone would like to be able to read it e.g. after 200 years a true digital archive should be used.

Conclusions and suggestions

Education in the field of digital preservation and information management is still at a very unstable level. First contact with the topics won't come until the students are 18 years-old and this in authors' opinion is way too late. Virtually all under 16-year-old children spend time online and whether they know it or not, they produce material. Some of this produced micro history might be valuable for future generations and therefore the creator should possess a basic knowledge of identifying and preserving possibly valuable content. Unfortunately most of these creators do not know anything about handling or preserving digital material.

Could the preservation be done by automatic bots that harvest the internet? Technically yes, but then there would be issues with legislation such as the Copyright act. The will and knowhow of preservation for future generations must come from the user side and this won't happen in the large scale until proper education reaches enough citizens, currently it doesn't. If the preservation education can be embedded in the second level teaching, it will reach around 35 000 general upper secondary students and 45 000 vocational upper secondary students in Finland every year. If all these students would have a possibility to participate into a preservation education, the amount knowhow also in the private and business sector would be greatly increased in long run. Naturally the knowledge gained from this basic education could not be compared against the university training but at least the students would have an appropriate basic knowledge from this important field.

When it comes to the question of utilized storage or preservation solution, it should not be like using a bazooka to kill a fly, instead correct tools and methods should be omitted into use. Balance between, requirements, benefits, costs and 'worth preservation ratio' must be resolved before the correct tools and methods can be identified. The author won't see any reasons why e.g. memory organizations should build their own digital archives or pay large amounts of money for utilizing a true digital archive, when the offerings by cloud storage services could be enough for them. It is very unlikely that companies such as Google or Microsoft will fail anytime soon. If this worries, then why not use two simultaneous clouds from different vendors.

Unfortunately, nothing happens in the large scale until the general knowhow of preservation is shared with the public. Few courses and two university level trainings are not enough to make a difference proper second level education is a way to start and the modification of the basic syllabus is the goal.

References

- [1] The Archives Act, 23.9.1994/831 (avail. <http://goo.gl/Nce8w>)
- [2] The Basic Education Act, 28.6.1998 (avail. <http://goo.gl/I9CwJ>)
- [3] Finnish national board of education, Education overview (avail. <http://goo.gl/NWjUV>)
- [4] EU kids online – final report (avail. <http://goo.gl/uBQVU>)
- [5] R.A. Ibata et al. "A vast, thin plane of corotating dwarf galaxies orbiting the Andromeda galaxy," *Jour. Nature*, 493, 2013.
- [6] E. Gibson, "Data on the internet is permanent after 20 minutes". *Infosecurity magazine* (21/4/11) (avail. <http://goo.gl/4qa6J>)

Author Biography

Anssi Jääskeläinen has an IT MSc. (2005) from Lappeenranta University of Technology and a PhD (2011) with a topic considering user experience. He has an extensive knowledge of telecommunication, it, user experience and usability. His current interest is in long term digital preservation and related UX.

Mirja Loponen studied system design in ATK Institute in Helsinki in the 1970s. Since 1973 she has been working in specialist and managerial positions in Tietotehdas and Honeywell Industrial Automation in Finland. She has been working with the Mikkeli University of Applied Sciences as database specialist and development manager from 2004. She is also the chairman of the Union for Municipal Archives in Finland.