

Citizen Archive as a Case Study in Personal Information Management, MyData and Digital Archiving

Mikko Lampi, Noora Talsi and Miia Kosonen,

South-Eastern Finland University of Applied Sciences, Patteristonkatu 3 D,
50100 Mikkeli, Finland
{mikko.lampi, noora.talsi, miia.kosonen}@xamk.fi

Abstract. This paper is a pragmatic case study into personal and community focused digital archiving. We introduce the development of the Citizen Archive as a case study, combined with multiple viewpoints based on personal digital archiving (PDA) and personal information management (PIM). We also take into consideration the contemporary research and development of MyData. We maintain there should be bottom-up designed digital archives based on open technologies and human-centered principles. This kind of archives should not be institutional or regulated but be based on the existing know-how and the tools and methods of professional archiving, extended with a human-centered approach and novel developments from the information technology domain.

Keywords: digital archiving, personal archiving, personal information management, MyData, information system design, human-centered design.

1 Introduction

Today, we live in an information society where everyone has a digital footprint and probably a digital shadow as well. In addition, we purposely create and consume digital information. Some information is useful in our daily lives, some we call memories – both are valuable with different metrics. This information can be documents, photos, emails, tweets, audio, video, blog posts, digital receipts, wearables data, activity logs, purchase history, and whatnot.

The amount of data and information is increasing at an accelerating speed. Information overload is not a far-fetched description for the phenomenon. We need tools and methods to manage our lives and the information we deem valuable. In addition to understanding and managing the information, we want to make sure we do not lose these materials until we wish to do so. Information systems and the underlying processes are not necessarily designed bearing long-term access and preservation in mind. The temporal scope can range greatly from a few months to years or even from decades to generations.

Historically, typical sources of personal information are print documents, letters, diaries, photos, and videos, of which many have been already digitized. However, in many cases, trustworthy storages for these digitized artefacts are lacking. They are preserved on a set of USB sticks or cloud services with terms and conditions that do

not guarantee the personal collection remains safe, under the user's own control, and preserved over a long time.

The situation gets even more complicated with modern digital media tools and social media accounts, which allow everyone to share the aspects of their life story easily. Digital images, mobile videos and online communication like emails or social-network exchanges on Facebook may form the current "family archives" which are scattered around the internet. People produce rich sets of digital collections, growing every day, and making every individual as a culture himself or herself [1]. These are not just sets of digital data but "a set of artifacts that has the potential to chronicle your life" [1]. Furthermore, personal data cannot be treated as publicly available data or some non-sensitive institutional information. It requires special attention to the privacy, ethics, and personal value.

The mandate of institutional archives, such as national archives, city archives, business archives, are limited to the domain and scope included in their mission. Considering their role and resources, it is a necessary decision. However, this kind of restrictive, policy-driven, top-down approach is not suitable for personal and community level archives.

We propose that there should be bottom-up designed digital archives based on open technologies and human-centered principles. This kind of archives should not be institutional or regulated. However, they should be based on the existing knowhow and the tools and methods of professional archiving. This knowledge should be extended with a human-centered approach and novel developments in the information technology domain.

This paper is a pragmatic case study into personal and community focused digital archiving with multiple viewpoints based on personal digital archiving (PDA) and personal information management (PIM). Furthermore, we take into consideration the contemporary research and development of MyData. We explore the digital personal archiving domain and try to propose solutions to the shortcomings of the existing tools via a multidisciplinary approach. Furthermore, we try to identify some of the roles of personal digital archiving in the bigger scope in information governance, research, and institutional archiving.

The rest of the paper is organized as follows. The second chapter is a brief introduction into the Citizen Archive. It describes the system and its design philosophy, and its use as a use case throughout this paper. The third chapter discusses personal digital archives and especially our viewpoint of it as well as the needs, purposes, and shortcomings for PDA systems. The fourth chapter is about PIM and its connection with PDA, and existing research regarding this intersection. The fifth chapter introduces the concept of MyData and its impact on PIM and PDA. Finally, we conclude the paper with discussions, implications, and future ideas in chapter six.

2 The Citizen Archive as a use case

Citizen Archive (see Jääskeläinen & Uosukainen, this volume) is an open-source solution for personal archiving developed at Digitalia, in South-Eastern Finland

University of Applied Sciences, Finland. It is based on a professional archiving tool that is being modified to meet the needs of citizens.

As pointed out by Pimminger et al. [2], typically the solutions for managing digital life stories cover only one aspect of the whole. For instance, they support the backup of information or sharing folders with other users such as family members. There is a lack of personal archiving systems that would support the whole process of archiving content, enabling reliable long-term preservation, and allowing users to easily manage (e.g. search, share, present or delete) their collections. This is the aim of the Citizen Archive. In line with MyData principles, users may decide what to do with their collections. User rights can be given to other users, family members, researchers, institutions or even firms.

We could identify four different approaches to the Citizen Archive. From the technical perspective, it supports preserving various types of personal information and linking the related content. The materials are pretty much static but can accumulate more metadata, either contextual or descriptive. Furthermore, they are organized in a way that is easy to navigate and to understand the context.

From the human perspective, the service is developed in collaboration with end-users and pilot customers, following the principles of user-centered design [3] and co-creation, where the role of the user/consumer has changed “from isolated to connected, from unaware to informed, from passive to active” [4].

Thirdly, for communities such as families the archive provides a community memory, which allows people to collect and represent their valuable digital heritage and make sure it is maintained over generations – in case the personal archivist has made a will covering his or her collections. An agreement between the user and the service provider is not enough to guarantee the long-term preservation.

Finally, the Citizen Archive also has its broader impact on society, above all by providing collections that official archival institutions are not interested of. As information and content are increasingly produced by large masses of people (e.g., citizen journalism, open source projects, personal video channels, blogosphere and so on), respectively there is a growing need to be able to preserve that content to the future generations of historians. The motivation for developing the Civil Archive include strengthening the civil society, digital rights, and transparency. The citizens own their data and it should be possible for them to provide and manage the access to the data and to govern how it is used.

3 Personal digital archiving aspect

As described in the Introduction, having just personal data storages is not enough to guarantee that our life stories remain safe, preserved, and usable. There is a need for solutions that cover a wider variety of needs than just storing information. Personal digital archiving needs governance too, yet being different than with professional archives.

Marshall [5] has identified four major challenges in personal digital archiving. Accumulation of data makes classification and value determination difficult. Amount of digital services can create silos and make management of distribution hard or

deceptively easy. Curation, including managing files, metadata, and migration, is something that the institutional archives are good at but individuals are often not. Finally, long-term access can be difficult to ensure with multiple online services and offline copies. Do people remember what they have saved and where? How long these services or locations are available?

The Citizen Archive tries to cope with the above challenges. It is a professional tool made for individuals. It takes care of distribution, curation, and access, but does not take away the decision-making power. It even provides some tools for classification. However, at the time being, the value determination remains a human task.

The research area of personal digital archiving is over the years trying to understand the nature of personal records; understand how people think about these collections, what meanings and importance they assign to them, and how these views change over time. Also, it is important to study not only what people are doing with personal data they create today, but also how they use it, if they share it with others and what has happened to the data collections that they created decades ago. [6].

Lynch [6] writes that the private becoming public is the most poorly understood and yet perhaps the most important area within the overall personal digital archiving research agenda. It is central to understanding the nature of personal digital archiving because it focuses on the connecting one and the other – private and the public. However, there is no requirement that personal digital archives or digital lives ever become public. Even though when private life is represented public more often in social media also archiving are increasingly public.

The previous studies have found several reasons on why people are archiving. The main reasons and the meanings are the same that have been found of using Citizen Archive. In the article of Kaye et al. [7], these reasons are for both digital and manual archiving.

- 1) Finding it later. One obvious reason people archive materials is that they can easily find it later. Finding and managing digital information is constantly creating new personal archiving tools. Also, different services as Dropbox, Google, Microsoft OneDrive etc. are answering the increasing need for organizing and archiving digital data from different sources. The Citizen Archive is more than just an online file storage. Materials in an archive are meant to be preserved for a specific period or permanently. The usual file storage services online are not designed for preserving information and maintaining control over it. This has caused scattering of data and difficulties in coherent and user-focused information management. We have created the Citizen Archive as one answer to this.

- 2) Fears of Loss. Another purpose for personal archiving is simply to keep data safe. Fearing of losing economically, socially, or emotionally important data, is as important as finding and managing the data. Many SaaS model based solutions are technically safe but vulnerable to unilateral changing of terms, conditions, and privacy policies. There is no guarantee that the data is stored for years to come. The Citizen Archive is solving this problem by offering safe environment with servers located in Finland (the same country the service is provided to), and a professional tape drive replication and backup.

- 3) Sharing Resources. Instead of keeping data of one individual alone, the Citizen Archive support the necessary retrieval of documents by many people. Kaye

et al. [7] have identified sharing as an archival goal common to many of the personal archives. Sharing the data among relatives, research community, organization or wider public is also the goal of Citizen Archive.

4) Building a Legacy. Building a legacy and creating a testament of life work are the essential reason for archiving personal data. Being remembered is the reason many citizens are keeping and managing personal data. The data could be willed for relatives, research purposes or just in case someone would find it interesting. [7].

Above all mentioned archiving personal data be an identity construction process. As Goffman [8] describes in *The Presentation of Self in Everyday Life* people are representing themselves through their practices and artefacts around them. The data that is valuable and meaningful enough to archive become a part of the constructed identity of a person.

Considering personal digital archiving as a self-constructing process or a way of building a legacy, the Citizen Archive has been working on another unsolved issue: death. Personal archive can be a digital legacy, and it can belong to a community rather than an individual. For example, a family archive could contain multiple personal archives. A personal data archive could include materials which terms of use do not allow inheritance. Moreover, often personal archives are connected to the context of a community and they cannot be understood in isolation [6]. The community might be dynamic and complex.

This is also the area where PDA is connected to a bigger whole, the national or even an international memory and cultural heritage. It can enrich and invigorate materials in institutional archives, especially when personal archives are opened to the public and connected to each other. The network of personal archives can create collective digital memory of ordinary citizens and mundane data. This is also why the Citizen Archive development has raised interest across a variety of communities and their representatives: it could assist in preserving the history of the “unseen” and “unheard”, that would cease to exist if relying only on archival institutions.

4 Personal information management aspect

Converging the approach in this paper, PIM is a multidisciplinary domain consisting of human-computer interaction (HCI), information retrieval, information science, cognitive psychology, and such. We primarily focus on PIM's connection with personal digital archiving in the light of previous studies. There is existing research (see e.g. [7], [9], and [10]) on the connection. Also, some research regarding personal digital archiving takes approach like PIM (see e.g. [11], and [5]).

Typically, PIM technologies cover three primary functions: creating, arranging, and accessing personal information. Archiving can be considered as a fourth function. Some archivists (e.g. [9]) link personal digital archiving strongly to archival science and the continuum model. Bass [10] argues that principles of provenance, trust, and availability over time should be adopted from professional archiving into PIM. Furthermore, value identification is a crucial activity in both PIM and PDA. We agree with these arguments and it is the rationale behind the Citizen Archive's objective to make professional archiving tools suitable for everyday use in personal digital

archiving. The focus in the Civil Archive is on the arranging, accessing, and archiving information. We also see PDA as a combination of both PIM and the traditional PDA activities. A personal digital archive contains materials which are preserved but can also be used for various purposes, re-organized or perhaps worked on. In the Citizen Archive, users may organize, browse, search, share, present and enrich their materials. There is no end of life-cycle in all the personal materials, or even need to define a life-cycle.

Henderson [12] identified three basic strategies people use to archive information. Filing strategy emphasis organic construction of a directory structure to support sorting and classification of materials. Browsing is the preferred access method. Piling strategy results in somewhat unorganized clusters of materials, usually in computer desktop or a similar storage. The emphasis is on the ease of browsing and being able to visualize the materials. In contrast to piling, search is an important feature to find and locate information. Structuring strategy uses a fixed and pre-designed hierarchy of directories. Materials are then classified and placed into the hierarchy.

Barreau & Nardi [13] argue that people do not often create, and thus use, a classification system. Especially, if it takes a lot of effort to plan and adhere to the system. Instead, people might use location as an aid to determine the context of the findings or use a text-based search tool. It is therefore necessary that a personal digital archive supports multiple strategies and methods for organizing and accessing the information. It should guide towards the best practices and teach good archiving but not restrict, limit, or demand a lot of preliminary planning.

People accumulate a lot of information over time. PIM approach to manage this growth is often a periodical loss of information. People tend to postpone the decision of preservation and disposal [11]. We feel that the information system, whether archival system or an information management system, should help in the decision making and value appraisal. The difference to the institutional archiving is that there are no policies on what to keep.

Lifestreams (see [14]) project took an interesting approach on managing the information flows. There are no hierarchies but documents are born into a timeline. The documents can be classified and browsed based on their context instead of location in a predetermined hierarchy. This is a similar approach as in the facets or virtual metadata-based views in the Citizen Archive. Using metadata to create a new viewpoint into personal information system or an archive is a tool for finding but also exploration.

Considering the three above findings and experiments is essential in developing intuitive tools for personal archives. The Civil Archive is aiming to support various strategies instead of defining a fixed strategy. A classification system can be used but it is not mandatory. Instead documents can be explored and accessed based on their contextual metadata.

5 MyData - a human-centered approach into personal data

The MyData whitepaper by Poikola et al. [15] defines MyData as a change in thinking in personal data management and processing. The shift is from organization centric systems towards human centric systems. Referring to the whitepaper, MyData has three core principles: 1) human centric control and privacy, 2) usable data, and 3) open environment.

It worth mentioning that the principles concern control and infrastructure instead of the actual ownership of personal data. Human centricity means transparency regarding the collection and use of personal data. In addition, people need to have the rights and practical means to control the data flows. This would then transform people from passive subjects into active, empowered actors. The practical means to control the data flows underline the usability of data. Usable data should be as up to date as reasonable, machine-readable, and equipped with adequate metadata. Online data should be accessible via standard APIs. Regardless of being online or offline, the data should be available in open file formats. Finally, the subset of personal data that complies with the MyData principles is also called MyData. Thus, the term means both the definition and the actual data.

MyData transforms data in closed silos into a reusable and crucial resource for the information economy. The value of data is generated through its circulation. However, it should be with the consent of the individuals. Data circulation requires both interoperability and portability. The MyData ecosystem is by nature distributed, and based on operators that can be switched without vendor locks. These operators store the consents for data flows. However, the actual data might not go through the operators.

The rationale for MyData can be inspected from the personal, business, or societal point of view, which are all important aspects. The personal point of view includes those of small communities, such as families, as well. The largest benefit is the control of data flows via the consents. A consent can be limited to a period, a specific purpose, or specific people or organizations. Control and transparency increase privacy and security. It is up to the users to decide if they would monetize their archive, allow research access, or keep the contents for personal purposes only. Perhaps, personal digital archive would be useful for personal analytics, or it could be used to enhance other services.

Businesses in the other hand, could rely on the fact that people, should they choose so, can provide them access to some of their data. Consents add transparency and trust, and the businesses do not need to gather the data from a scratch. Companies could be able to reach economies of scale benefits faster, acquire insights and co-create value. It sounds like a win-win scenario. However, there is a first-mover problem and reasonable fear of losing assets and competitive edge. It might be that national legislation or legislation by the European Union will provide the initial push. EU General Data Protection Regulation (GDPR) might be one the major drivers for MyData adoption. Poikola et al. [15] propose MyData framework as a mean to be ready for the GDPR.

Information society needs a data framework for its own digital services. Some of these services are digital personal accounts which could provide personal digital

archives, at least for some official records. This kind of framework could empower citizens and enable economic growth as well.

Trust via consent is a building block and a result of MyData infrastructure. If we look at the archives, this sounds very familiar. Personal digital archives could be more than just trustworthy repositories, they could act as personal data storage systems in the MyData ecosystem. Digital archives save provenance, enforce trustworthiness, and focus on preservation and access per se. MyData infrastructure adds in transparency, consent management and the ability to put the data into use. Digital archives could be services within MyData ecosystem but also a crucial component for preserving the consents. Therefore, digital archives would benefit from adopting the MyData viewpoint and vice versa.

MyData has gained increasing exposure in academia, business, and society for a few years. In Finland, for example, it has been included in some public funding programs for research and business, and government initiatives. There are also initiatives in other countries such as Midata in UK. Some research domains, such as PIM, HCI and even information governance, have shown interest in MyData or consent-based personal data management. We are preparing digital archiving to be ready to move into the human-centered MyData era.

6 Discussion

This paper has represented a pragmatic case study into personal and community focused digital archiving with multiple viewpoints based on personal digital archiving (PDA) and personal information management (PIM) as well as the contemporary research and development of MyData. We have explored the digital personal archiving domain and try to propose solutions to the shortcomings of the existing tools via a multidisciplinary approach. Furthermore, we have identified some of the roles of personal digital archiving in the bigger scope in information governance, research, and institutional archiving. We have proposed that there should be bottom-up designed digital archives based on open technologies and human-centered principles and presented Citizen Archive as an example. This kind of archives should not be institutional or regulated, but be based on the existing knowhow and the tools and methods of professional archiving.

As Marshall et al. [11] argue, special heuristics could be developed for value determination. In the future, perhaps the emerging artificial intelligence could become “a virtual archivist”. Also, the lessons learnt from the Citizen Archive development imply that personal archiving requires a lot of preliminary work, planning, and organizing, which is often the most laborious part for users. They clearly would benefit from virtual assistance in both preparing and compiling their personal collections.

Another stream for future research and development endeavors is to adjust to the various representations of information, not restricted to the record-based view by archival institutions. For instance, in online network exchanges, an information object is not a single message but a set of messages and the connections between their senders. How to archive social interaction? There is yet a temporary, shared context

that links the information together. From PDA point of view, we do not have only one personal digital identity but multiple ones, as each of us represents himself or herself a bit differently regarding which platform we are using: Facebook or WhatsApp chats with family and friends, emails for official tasks, or Twitter for public discussions with a wider audience. Hence, in the future it is of essence to be able to gather personal information on a cross-platform basis from different online sources and services to the Citizen Archive and archive the socially represented digital self.

Some of the reasons why people archive is connected to the need to manage personal information while others are more about preserving information. We argue that personal digital archiving is indeed a hybrid domain. Therefore, the archiving solutions should be aligned with it. We also maintain that developing “citizen archives” would benefit from transcending silos and combining skills and resources across a variety of fields, of which archiving policies, digital archiving, personal information management and MyData were highlighted in this paper.

A common use case in personal archiving starts by the creation of a personal archive and ends when it is donated to an archive [9]. The perspective is linked with records life-cycle thinking, and Cushing argues that professional archiving should act earlier. We agree that people probably lack interest for time consuming stewardship and systematic records keeping. But the solution could be automating the tedious tasks instead of moving away from the people. It is an interesting question whether we should empower the people, as in MyData thinking, or encourage archival interventions. After all, information management and digital literacy are must have skills in the information society and could increase social transparency and democracy.

References

1. Carroll, E. & Romano, J.: *Your Digital Afterlife. When Facebook, Flickr and Twitter Are Your Estate, What's Your Legacy?* New Riders, Berkeley, CA (2011).
2. Pimminger, S., Heinzlreiter, J., Kurschl, W. & Lindley, A.: *Themis - Conserve Your Digital Life*. FFH 2015, Forschungsforum der österreichischen Fachhochschulen, Hagenberg (2015)
3. Lowdermilk, T.: *User-Centered Design a Developer's Guide to Building User-Friendly Applications*. O'Reilly Media (2013)
4. Prahalad, C. & Ramaswamy, V.: *Co-creating unique value with customers*. *Strategy & Leadership*, 32(3), pp. 4-9 (2004)
5. Marshall, C.: *Rethinking Personal Digital Archiving, Part 1*. *D-Lib Magazine* 14(3/4) (2008)
6. Lynch, C.: *The Future of Personal Digital Archiving: Defining the Research Agendas*. *Personal Archiving: Preserving Our Digital Heritage*. *Information Today* (2013)
7. Kaye, J., Vertesi, J., Avery, S., Dafoe, A., David S., Onaga, L, Rosero, I & Pinch, T.: *To have and to hold: exploring the personal archive*. CHI (2006)
8. Goffman, E.: *The Presentation of Self in Everyday Life*. Penguin (1959)
9. Cushing, A.: *Highlighting the archives perspective in the personal digital archiving*. *Library Hi Tech*, Vol. 28 Issue: 2, pp. 301-312 (2010)
10. Bass, J. *A PIM perspective: Leveraging Personal Information Management Research in the Archiving of Personal Digital Records*. *Archivaria* 75, pp. 49-76 (2013)
11. Marshall, C., Bly, S. & Brun-Cottan, V.: *The Long Term Fate of Our Digital Belongings: Toward a Service Model for Personal Archives*. *CoRR*, abs/0704.3653 (2007)

12. Henderson, S.: Personal document management strategies. CHINZ (2009)
13. Barreau, D. & Nardi, B.: Finding and reminding: file organization from the desktop. SGCH (1995)
14. Freeman, E. & Gelernter, D.: Lifestreams: A Storage Model for Personal Data. SIGMOD Record, 25, pp. 80-86 (1996)
15. Poikola, A., Kuikkaniemi, K. & Honko, H.: MyData - A Nordic Model for human-centered personal data management and processing. Ministry of Transport and Communications (2015)