

# Gluing Provenance to Dispersed Personal Content and Creating Contemporary Personal Archives

Arian Rajh

Agency for Medicinal Products and Medical Devices /  
Department of Information and Communication Sciences,  
Faculty of Humanities and Social Sciences, University of Zagreb /  
Highflott  
Zagreb, Croatia  
arian.rajh@halmed.hr

Krešimir Meze

Think Big Hub  
Zvonimira Furtingera 1, Zagreb, Croatia  
kresimir.meze@thinkbighub.com

## Summary

*The article discusses the necessity of providing archival support to creators of personal archives. The authors state the need to overcome the fragmentation of contemporary personal fonds, bind together overall personal content and facilitate long-term preservation. The article focuses on the preliminary communication of personal digital archiving requirements and presents the development of the prototype software product.*

**Key words:** Personal Digital Archiving (PDA), digital legacy, prototyping

## Introduction

Personal archives or fonds are entireties of records created and accumulated by individuals during their lives. They appear to have distinct qualities compared to other types of archives because the transactional value of the records is not their prevalent value (Hobbs 2001, 128). According to Hobbs, personal archives are “archives of character” and they are “documenting our complex inner humanity“ (Hobbs 2001, 135). The same definition should also relate to personal digital archives.

“[T]hey accumulate as the person goes about his or her own private work and life’s activities and are ordered (or not) to suit the individual’s proclivities and needs.” (Eastwood 2016, 19). Contemporary digital personal archives,<sup>1</sup> in the

---

<sup>1</sup> “The term ‘personal digital archiving’ refers to how individuals manage or keep track of their digital files, where they store them, and how these files are described and organised. People keep personal archives for reasons that may be simultaneously sentimental, practical, and even accidental.”, Redwine, G. *ibid.* Furthermore, “There is an emerging scholarly discipline, personal in-

majority of instances, will hardly end in public archives where they would be arranged and kept according to the professional standards. Even with the best will in the world, the accumulation of digital personal records and content surpasses the capacities of professional processing in archival institutions. Personal archives will not be structured and arranged by archivists and there is a considerable risk of loss of their content. Today it is recognised that public institutions are creating just a part of future memory and that private records creators accumulate socially and culturally relevant archival materials too. This adds severity to the risk of continuously losing potentially valuable archival holdings.<sup>2</sup>

Personal archives have particular ingest policy, organisation, structure and a notion of value. They are also different from institutional papers concerning the transfer to archival institutions and their processing. Maybe it's time to change our approach to personal archives – if they are different from the typical archives that archival professionals process in the archival institutions, and if their accessions to archival institutions are not likely to happen. Preserving potentially valuable archival holdings should be the main drive for archivists to intervene and assist individuals in their preservation efforts. “The archival profession needs to concentrate on developing new mechanisms for education the public about how to care for their personal and family archives” (Cox 2009, 107). Because of large production of personal materials, we propose an additional way to do it – with a help of ICT tools with built-in professional knowledge. This proposal is in line with contemporary archival mission. There are a number of tools for content creation, capturing, storing, sharing and recycling available to individuals. However, there is a lack of tools which might ensure content cohesion - tools for gluing records together, gluing provenance to the content and for organizing records and saving their interpretative potential for the future. Provenance refers to the archival principle of separating material of different origins (and grouping material of the same origin).

Archival tools for the organisation of personal fonds are unavailable to personal records creators (end users). This leaves personal accumulations unstructured,

---

formation management, archivists need to begin both to dig into and to influence (and there is probably a natural connection to the other emerging area, digital curation).”, Cox, R. J. *Digital Curation and the Citizen Archivist*. //: *Digital Curation: Practice, Promises & Prospects*. Uni. of North Carolina, 2009, p. 108.

<sup>2</sup> More creators archive their records than in the past and creators of personal archives are included in social power dynamics. Their small narratives and records could become important in the future for seeing the bigger social picture. That is the reason why archival profession should proactively offer methodologies and tools for assembling of personal archives. If personal archives are better organised and arranged, it will be easier to use the legacy of their creators in the case of acquisitions in the future. After all, “the more control an individual has over their personal archives, the greater their ability to save only what they intend”, Redwine, Gabriella. *Personal Digital Archiving*. DCP Technology Watch Report 15-01, December 2015, Digital Preservation Coalition, <http://dx.doi.org/10.7207/twr15-01> (accessed January 2017), p.2.

provenance links decaying and content scattered. Our research questions are linked to the quality of emerging tools for personal digital archiving (PDA) and to characteristics of personal archives:

1. *How to create personal digital archives today?* More exactly, what kind of user and professional archival requirements should PDA tools support in order to facilitate personal digital archiving and to mitigate the risk of losing personal content?
2. *What is the purpose of creating personal archives?* What characteristics should contemporary PDA tools and personal archives have in order to promote that purpose?

### **From PDA requirements to PDA model and prototype**

To answer the first set of research questions we analysed and divided user and professional requirements to those that address long-term preservation of content and those that address content accumulation. After requirements analysis, we proposed mechanisms for their implementation in two stages: on a conceptual level of model and on the functional level of actual prototype tool. We are at the stage of developing the prototype now and we are preparing it as open source software and commercial SaaS solution. Later in this article, we will show use cases which will be covered by our prototype and which should also, in our opinion, be covered by other potential ICT solutions.

In Table 1 we have listed the requirements that PDA software model and the prototype should satisfy. These requirements are central problems that PDA conceptual model and tool should solve; implementation mechanisms on the model level are functional solutions to the stated problems and implementation mechanisms on prototype level are software functions. We are using the methods from Table 1 for the development of the prototype.

The first content-related requirement (IA) deals with the usage of various cloud file hosting services and mail solutions. The personal digital archiving solution should be able to serve as the portal or central point of access to personal content stored in various environments (e.g. digital repositories, mail services, social media platforms, professional and scientific social networking sites, shared content). We imagine digital personal fonds<sup>3</sup> as portals – this is based on the idea of fonds as intellectual constructs or conceptual aggregations in archival theories of Barr, Cook and Yeo (Barr 1988; Cook 1993; Yeo 2012). Creators should be able to have access to their records and have them readable and usable over the long term. Creators' records include files of various file formats,

---

<sup>3</sup> Fonds – “[t]he entire body of records of an organization, family, or individual that have been created and accumulated”, Glossary of archival and records terminology, <https://www2.archivists.org/glossary/terms/f/fonds> (accessed 26/7/2017)

e.g. textual formats, image formats, multimedia container formats.<sup>4</sup> Content LTP requirement (IB) presumes proactive monitoring of the content. Although certain qualities of the content should be assured before archiving, in the act of creation of the content, a tool for personal digital archiving should monitor the targeted properties and enable the creator to act at the right time. PDA prototype was not intended to be used for conversion of file formats – our viewpoint is to use functionalities and tools already developed by other IT companies and to focus on archival issues.

Table 1: Methods used for functional model and prototype of proposed PDA solution

	Requirements	Model level implementation – concepts	Prototype level implementation – functions
<i>I</i>	<i>Requirements related to digital content</i>		
A	Provide access to personal content	Portal that enables a central access point for selection and processing of content for personal fonds	<i>Add account</i> function for the integration with various software solutions by using open web services
B	Content long-term preservation (LTP)	Relying on OAIS and related standards for monitoring all content and export of information packages	<i>Analyze&amp;report</i> function; <i>Export</i> function
<i>II</i>	<i>Requirements related to aggregation of digital content</i>		
A	Establish virtual provenance of fonds	Binding all selected content into the personal fonds by using the pre-set structure	<i>Select units for my fonds</i> function; <i>Create/update my fonds</i> function
B	Facilitate interpretative potential of fonds	Relying on archival professional standards for description, formatting standards, cognitive services for indexing, linked data and semantic web for selected descriptions	<i>Add archival description</i> function (with advanced features, e.g. linking with cognitive services for photos in the fonds); <i>Make selected description available</i> function

Besides these two *content-oriented requirements* (IA, IB), there are two requirements *linked to content accumulation* or fonds (IIA, IIB). Binding widespread content into personal fonds (IIA) is the central concept around which our PDA tool should be built. The tool should propose the structure of personal fonds to end user and the placement of the archived item in the structure. User modifications of the structure should be allowed because there are differences between various personal fonds and the final structure depends on many factors related to creator and material. The archival description should be added (IIB) for the purposes of taking intellectual control of the content, for retrieval and for

<sup>4</sup> The technical registry PRONOM, accessed 25/7/2017 <https://www.nationalarchives.gov.uk/aboutapps/PRONOM/default.htm>

preservation. It should be possible to track identity, integrity and captured context(s) of creator's records. We are considering using cognitive services could for additional content analysis and indexing. All of this brings us back to the second set of research questions which is about a purpose – or purposes – of personal archives. Records creator creates personal fonds to keep his or her records and contents together, to keep personal memories and evidence.<sup>5</sup> This purpose is expected. However, although personal archives are relatively isolated from the environment, archives, in general, show a tension between availability and unavailability. Openness for usage is, of course, closer concept to public records creators' holdings. But it is linked with any accumulation of memories. To some extent sharing of selected content of personal archives is similar to the writing of "personal" memoir writing or diaries.<sup>6</sup> Any writing is supposed to be read; not just by the author; it is something fused with its very existence. The perplexing situation with archived items of personal provenance resembles the situation with autobiographical prose; they want to be found. So, this secondary purpose is something linked with the desire to participate in wider social memory. Sometimes records creators want to be present in a society and want to contribute, promote him or herself, provide or add something related to his or her point of view. In the end, a local historian would be very glad to retrieve and to access photographs or other evidence of locally significant historical event taken and provided by private records creator. We have considered this second purpose also in the course of prototyping. That is why we have designed characteristics that contemporary personal archives need to enable participation in the shaping of the new archival landscape of tomorrow. It could also be convenient for accruals of material into an existing personal archive or for sharing archived items with family members who live apart (which could lead personal records creators to the creation of virtual family archives). Providing access to selected descriptions and then to photographs, videos and other records, could be facilitated by using linked data and semantic web concepts. Appropriate archival description assures interpretative potential of the fonds and enables preservation of personal heritage for future use by the creators or his or her heirs. It's like sharing action... but with archival qualities.

---

<sup>5</sup> "With the proliferation of networked information and communication technologies (ICTs), the definition of personal archives has expanded to include the collections of potentially any individual with an archival impulse to document his or her life.", Acker, Amelia; Brubaker, Jed R. *Death, Memorialization, and Social Media: A Platform Perspective for Personal Archives*. // *Archivaria* 77 (2014), p.3.

<sup>6</sup> Please see the example taken from a contemporary website called Public Diaries. This service hosts private and public diaries. Many authors decide to make their diaries openly accessible, <https://www.my-diary.org/surf/> (accessed 26/7/2017).

Having both sets of research questions in mind, we decided to cover and automatize PDA use cases of *personal fonds creation* (creation and arrangement of archival information packages and archival units, creation of descriptions)

1. *intended participation in wider memory with selected materials*
2. *preparation for preservation of information packages with the export function*

We think that PDA tools, in general, should implement these fundamental use cases. The desirable contemporary PDA tool should be easy to use and it should reflect the smart integration of ICT and professional knowledge into daily life. Based on this analysis of requirements and these use cases, we<sup>7</sup> are creating the prototype of PDA solution called “Legacy Sky”<sup>8</sup> which will enable building personal fonds and preservation of personal digital legacy. PDA tools in general should solve the problem of dispersing storage and facilitate access to digital content stored on any digital repository. It should help users to preserve their digital content and targeted properties of the content. It should help users to organise their content. In one sentence, the tool should bind the content scattered through various spaces and add archival structure and description. From that point onwards, content takes the form of personal fonds. It can be preserved as a coherent unit from that point on. Storage and access to content, as well as preservation issues, are connected with requirements related to digital content and structuring and adding the description with requirements related to aggregation of digital content.

### **PDA prototype and its possible impact**

We strive to implement previously stated PDA requirements and develop PDA tool prototype. We assume that innovative linking of different cloud services (repositories, social networks etc.) into one archives will provide new value for personal records creators. For prototyping PDA software product we are using lean-based project management method that combines several different lean canvases as well as state-of-the-art *scrum* methodology.<sup>9</sup> The prototype is based on web portal connected with content storage providers through open web services. It is also based on several straightforward software functions which provide a solution for previously identified requirements.

---

<sup>7</sup> Highflott (<http://www.highflott.com>) and Think Big Hub (<http://www.thinkbighub.com/>).

<sup>8</sup> <http://www.legacysky.com>

<sup>9</sup> Business model canvas by Alexander Osterwalder and Yves Pigneur; lean canvas for start-ups, adapted by Ash Maurya. The canvas was used for developing Legacy Sky start-up as the joint venture between our companies. For the development of our prototype we are using *scrum* software development methodology (GitHub Boards, Agile project management <https://www.zenhub.com/>). We plan to include early adopters in our *scrum* development process.

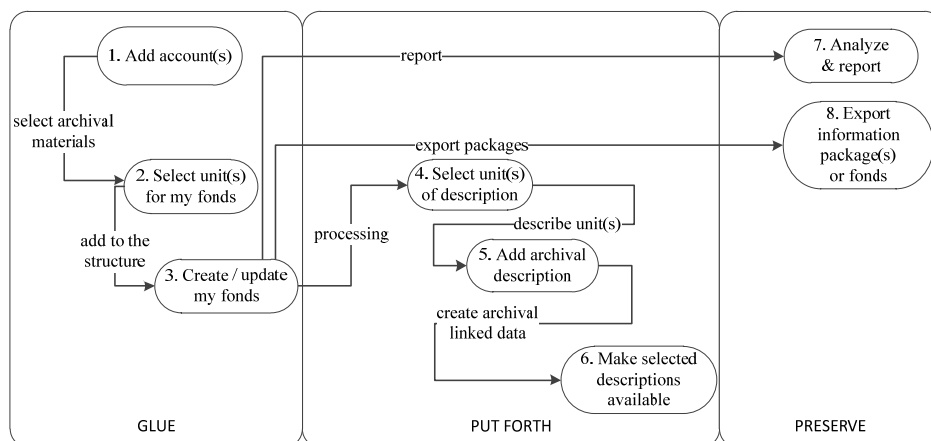


Figure 1: High-level functional model of the PDA tool Legacy Sky prototype

Archival functions of the tool that “glue together” personal content are: (1) add cloud service account and (2) select units for fonds. The functions in control for description and (consequent) visibility of archival materials of personal fonds are: (3) create/update fonds, (4) select units for description, (5) add description (including metadata created by cognitive services for the description of images) and (6) make selected descriptions (and units) available. The functions that support the proactive preservation of archives are related to (7) analysis of file formats and (8) export of information packages (in OAIS sense) and they are still work in progress. The tool should propose an archival structure of the personal fonds for the organisation of personal content<sup>10</sup> and lead end users to add the archival description in an easy manner. The offered generic structure should facilitate placement of records by the creators and facilitate growth and maintenance of personal fonds. Although generic structure should be suggested by PDA tools, they should support certain flexibility during the development of specific structures, especially knowing that this flexibility is quite common in the archival processing of personal fonds. Advanced users should be able to adjust the structure so that it could match their *character* better (Hobbs 2001) or

<sup>10</sup> E.g. personal records, correspondence, notes, records created in business activities, auxiliary records in Škalić-Štambuk, M. Vrednovanje arhivskog gradiva u osobnim arhivskim fondovima. // Arhivski vjesnik 38 (1995), 88-89. Another generic structure – biographical and personal records, correspondence (except messages inseparable from particular professional business activity records, cooperation or project records), professional and other activities, the body of work/œuvre, ownership and financial records, testimonies of others, collections, varia (with sub-series formed by creators). Lucic (2014) mentioned several possible structures for personal archival fonds and family fonds. But, there is no one “perfect” structure, so we are “stuck” with generic structures that could be applied for the majority of archives.

*present the life and work of creator*<sup>11</sup> in a more apparent way. Users should be able to add subseries for sorting out closed documentary units. Subseries of business-related records, cooperation and projects series should be harmonised with creator's activities. Development and usage of such IT tool would provide many users with a helpful possibility to preserve their content recorded in various repositories for the future. Innovative linking of various online services by using such a tool is going to create additional value in PDA domain. If tools are built on professional concepts and methods, individuals who create and store their records and content will preserve them in the proper way. With creators participating in their archives' environments, this could have an impact on the preservation of future memory in a proactive way.

### **Conclusion**

The risk of losing memory relates to conventional and digital personal archives. M. Lučić showed in her book about personal archives that 43% of personal fonds *preserved* in Croatian cultural institutions consist of one box of material. This analysis was based on the example of 1949 personal archives preserved in Croatian National Archives and other cultural institutions in Croatia.<sup>12</sup> Analysis of completeness of personal archives showed that 7% of them relate to one year of creator's life and 5% of them relate to the period of 2-10 years.<sup>13</sup> Lučić stressed out a deficit of material and/or structural asymmetry of preserved personal archives. The principle of provenance was not respected when materials of the same creator were scattered in different physical institutional repositories and without linking (Lučić 2014, 64-66). We share Lučić's concern for personal archives and consider potential harm very critical due to extensive production and proliferation of digital materials, as well as their volatility. Usage of different repositories without having a portal that glues the pieces together is the same as having one particular creator's legacy in custody of various institutions. Everybody creates digital content, manages these materials through various repositories and experience obsolescence problems severe enough to require some action. "We are all archivists", as Richard J. Cox stated in one section heading of his 2009 article.<sup>14</sup>

---

<sup>11</sup> Lučić, M. *Osobni arhivski fondovi: Arhivistički pogled na prikupljanje, obradbu i interpretaciju rukopisnih ostavština u baštinskim institucijama*. Zagreb: HDA, 2014, p. 123.

<sup>12</sup> Lučić, *ibid.*, p. 57.

<sup>13</sup> This analysis was based on General guide for archival fonds and collections in the Republic of Croatia (2006) finding aid, Lučić, *ibid.*, p. 58.

<sup>14</sup> Cox, R. J. *Digital Curation and the Citizen Archivist. //: Digital Curation: Practice, Promises & Prospects*. Uni. of North Carolina, 2009, p. 107.



We can agree with Hobbs's perception of personal archives as different from public records and with Cox's perception of archivists' role<sup>15</sup> in their digital preservation. Postcustodial archivists ought to be advisors. The definition of archives should be broadened to include not just results of practical activities of institutions but also someone's personal legacy.<sup>16</sup> There are difficulties of processing personal archives by archivists in formal archival institutions and these difficulties are even more severe in a domain of personal *digital* archives. Archival processing done by the professionals in archival institutions cannot possibly meet the needs of all records creators. That is why we considered an approach based on the principle of helping people to help themselves. This approach comes down to creating new tools with built-in archival logic for records creators. Contemporary personal digital archives should be created by using current technologies and professional logic but in a creative manner and according to the needs of existing personal archives creators. They are quite different from 19th Century gubernatorial organisations. This was related to the first research question. The purpose of creating a personal archive, to answer the second question, lies in the preservation of content for any other personal purposes, including content sharing and thus participating in wider information realm.

PDA enables taking timely actions, it envisages new role of archival knowledge and ICT tools in the contemporary world, and it answers the requirement to support a new form of digital meta-literacy. At last, reinforcement of creation of personal archives by using archival logic already built in computer applications, so users are not burdened with professional principles, represents a possible further integration of ICT into people's daily life. What does it mean? A person doesn't have to be famous to have his or her personal archives arranged professionally and described properly. Capability to archive his or her own personal content in a form of fonds and even to offer archival material to the community represents archiving democratization (similar to data democratization). There is too much digital content nowadays and archiving should be facilitated by ICT tools with built-in archival logic. Users should be able to keep their digital legacy and even *participate* in the creation of the digital landscape of tomorrow.

---

<sup>15</sup> "In this role archivists will function more as advisors rather than acquirers, educators giving their knowledge away rather than protecting the secrets of a guild, and advocates rather than reactors in seeking to preserve the portion of the documentary universe that possesses archival value.", Cox, *ibid.*, p. 108.

<sup>16</sup> Please see Melina Lučić's elaboration of a sharp distinction between public archives and personal archives (or legacy) in more traditional archival theories (Jenkinson, Casanova, Brenneke, Duranti) and merging of these notions to one definition of archives in more recent or more flexible theoretical approaches (Lučić 2014, 16-21).

## References

- Acker, A.; Brubaker, J. R. Death, Memorialization, and Social Media: A Platform Perspective for Personal Archives. //: *Archivaria* 77 (2014), 1-23.
- Barr, D. The Fonds Concept in the Working Group on Archival Descriptive Standards Report. *Archivaria* 25: includes supplement, "The Archival Legacy of the Department of the Interior" (Winter 1987-88)
- Barrett, H. Personal archives: Documenting the stories of our lives. 27.6.2014. Financial Times. <https://www.ft.com/content/b1bf5044-f5ba-11e3-afd3-00144feabdc0> (February 2017)
- Cook, T. The Concept of the Archival Fonds in the Post-Custodial Era: Theory, Problems and Solutions. *Archivaria* 35 (1993), pp. 24-37.
- Cox, R. J. Digital Curation and the Citizen Archivist. //: *Digital Curation: Practice, Promises & Prospects*. University of North Carolina School of Information and Library Science, 2009, pp. 102-109.
- Cox, R. J. *Personal Archives and a New Archival Calling: Readings, Reflections and Ruminations*. Duluth, Minn.: Litwin Books, 2008
- Eastwood, T. A Contested Realm: The Nature of Archives and the Orientation of Archival Science. U: MacNeil, H.; Eastwood, T. (ur.). *Currents of Archival Thinking*, 2. izd., Santa Barbara: Libraries Unlimited, 2016., pp. 3-23.
- Guide to preserving your digital legacy. February 2015. Saga legal services. <https://www.saga.co.uk/saga/media/Legal/Digital%20legacy%20guide%20Feb%2015/Digital%20Legacy%20guide.pdf> (February 2017)
- Gränström, C. Access to current records and archives, as a tool of democracy, transparency and openness of the government administration, *Arh. vjesn.*, god. 42 (1999), pp. 79-92
- Hobbs, C. The Character of Personal Archives: Reflections on the Value of Records of Individuals. // *Archivaria* 52 (2001), 126-135.
- ICA. Records in Contexts CM. Consultation Draft v.01, September 2016
- ICA. ISAD(G): General International Standard Archival Description, 2nd ed. 2000
- InterPARES project. <http://www.interpares.org/> (February 2017)
- ISO 14721:2012 Space Data and Information Transfer Systems – Open Archival Information System – Reference Model
- ISO 20104:2015 Space data and information transfer systems – Producer-Archive Interface Specification (PAIS)
- Lučić, M. *Osobni arhivski fondovi: Arhivistički pogled na prikupljanje, obradbu i interpretaciju rukopisnih ostavština u baštinskim institucijama*. Zagreb: HDA, 2014.
- Maurya, A. *Running Lean: Iterate From Plan A to a Plan That Works*, 2012.
- MacNeil, H.; Eastwood, T. (ur.). *Currents of Archival Thinking*, 2. izd., Santa Barbara: Libraries Unlimited, 2016.
- Muller, S.; Feith, J.A.; Fruin, R. *Manual for the arrangement and description of archives: drawn up by direction of the Netherlands Association of Archivists*. Chicago, IL : Society of American Archivists, 2003. <http://hdl.handle.net/2027/mdp.39015057022447> (July 2017)
- Osterwalder, A.; Pigneur, Y. *Business Model Generation: A Handbook for visionaries, Game Changers, and Challengers*, 2010
- Redwine, G. Personal Digital Archiving. DCP Technology Watch Report 15-01, 2015, Digital Preservation Coalition, <http://dx.doi.org/10.7207/twr15-01> (accessed January 2017)
- Škalić-Štambuk, M. Vrednovanje arhivskog gradiva u osobnim arhivskim fondovima. // *Arhivski vjesnik* 38 (1995), 81-91.
- Yeo, G. The Conceptual Fonds and the Physical Collection. *Archivaria* 73 (2012), 43-80.