

---

# Personal Photo Preservation for the Smartphone Generation

**Maria K Wolters**

Human Cognitive Neuroscience  
University of Edinburgh  
Edinburgh, UK  
maria.wolters@ed.ac.uk

**Francesco Gallo**

EURIX Research  
and Innovation  
Torino, IT  
gallo@eurix.it

**Elaine Niven**

Human Cognitive Neuroscience  
University of Edinburgh  
Edinburgh, UK  
e.h.niven@ed.ac.uk

**Heiko Maus**

German Research Institute  
for Artificial Intelligence  
Kaiserslautern, DE  
maus@dfki.de

**Mari Runardotter**

Information Systems  
Lulea Technical University  
Lulea, SE  
Mari.Runardotter@ltu.se

**Robert H Logie**

Human Cognitive Neuroscience  
University of Edinburgh  
Edinburgh, UK  
r.h.logie@ed.ac.uk

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

*CHI'15 Extended Abstracts*, Apr 18-23, 2015, Seoul, Republic of Korea  
ACM 978-1-4503-3146-3/15/04.

<http://dx.doi.org/10.1145/2702613.2732793>

**Abstract**

Preserving photos for future generations is difficult in the digital age, as both storage media and storage formats become obsolete within decades. In order to inform the design of a photo preservation service, we are currently collecting information about relevant practices in a large survey. In this paper, we report intermediate results from a sample of 236 European students aged between 18 and 34. 76% of our participants are keen to preserve their photos for future generations, but far fewer report photo management practices that support preservation. We discuss implications for design and outline three groups of users that can be distilled into personas.

**Author Keywords**

preservation; smartphones; digital photos; digital heirlooms; personas

**ACM Classification Keywords**

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; E.5 [Files]: Organization/Structure.

**Motivation**

With the advent of digital cameras and smartphones, the number of personal photos taken has increased exponentially. Many HCI researchers have studied issues related to photo management, search, and

retrieval [1, 7, 12]. Photos can also serve as personal mementos that allow people to revisit the past [4, 18], augmenting the physical mementos that are still at the core of personal heritage [14, 13].

Preserving personal digital mementos for future generations is hard [10]. Storage media and storage formats rapidly become obsolete, and backup storage media can fail [15]. In the ForgetIT project [5], we are developing a new approach to personal digital preservation that is based on three foundations: synergetic preservation, contextualised remembering, and managed forgetting.

*Synergetic preservation* means that digital artifacts to be preserved are regularly checked, and moved to new storage media if needed. This goes well beyond cloud storage services such as Google Drive, which primarily provide redundant storage. Through *contextualised remembering*, we ensure that digital artifacts are linked to the information required to interpret them. *Managed forgetting* recognises that not all digital artifacts are equally important. Some artifacts can be compressed, groups of artifacts can be summarised, and artifacts that are no longer important can be deleted or removed from preservation.

To support the design of our personal preservation system, we are conducting a large-scale survey of people from different cultures and age groups. Here, we present first results from a subset of data that focuses on young people and students who are mostly from Europe. We wanted to know how our respondents manage their photos, whether they are systematically leveraging the manifold opportunities to add context, and how they preserve what is important to them. We also used cluster analysis to determine whether there might be patterns of

photo preservation and management practices that could form the basis of personas.

## The Survey

Ethical Approval was granted by the Psychology Research Ethics Committee, University of Edinburgh.

The survey was designed using a three-step process, which is documented in [6]. Based on a review of the literature, we designed a pilot survey with several free text fields and distributed it across the consortium. Based on a content analysis of the free text fields and feedback from survey participants, we created a shorter, more focused version of the survey that could be completed in under 15 minutes and piloted it with 72 participants in a research study. The final version of the survey was implemented using LimeSurvey and distributed to students, via crowdsourcing, on the project web page, and through snowball sampling.

The survey consists of three main parts that cover sources and content of digital photos, photo management practices, and photo preservation practices. Demographic questions include a short instrument for assessing attitudes to technology developed for this survey and a modified version of Buchanan et al.'s [3] instrument for assessing online privacy concerns.

## Statistical Analysis

All percentages are rounded to full percent because the total sample size is below 500. Due to this rounding, percentages may not always add up to 100. Links between variables were tested using the Fisher Test as implemented in R. For cluster analysis, the k-means algorithm was used [9]. We explored cluster solutions ranging from 2 to 10 clusters and chose the solution that proved most

**Sample scenario:** Consider Kim, an HCI researcher who was at CHI 1997 and CHI 2010, both in Atlanta. At CHI 1997, Kim took digital photos with the brand new Sony Mavica; all photos were stored on a 3.5" floppy disk. For CHI 2010, Kim switched to the iPhone 3GS. Kim copied the content of the Mavica's floppy disks onto a hard drive, checked the drive regularly, and moved to a new backup drive every 2-3 years. The Mavica photos lacked meta-data, and Kim has forgotten many of the names and places, but the 2010 photos are tagged with date and location, which allows Kim to group them according to session (and party). While Kim still has some of the 1997 party photos on an old hard drive, they are no longer on the current photo backup disk.

stable, judging by the four stability criteria provided by the `clValid` package [2]. The coherence of each cluster is assessed using average silhouette width [16], where 1 indicates high coherence and -1 high dissociation.

## Results

### Participants

In total, 236 participants were recruited from German, Swedish, Italian, and British (UK) universities through the partners of the ForgetIT project. The German and Italian participants were predominantly male STEM students (Physics and Computer Science), the Swedish students were recruited from the general student population of Lulea Technical University, and the British students were first year Psychology students who were pointed to the survey as part of their course.

As Table 1 shows, the actual distribution of countries of origin is far more diverse than the sampling strategy suggests; this is mostly due to students who filled in the British survey. Almost all of the older students were Swedish. In most regions, respondents tended to be female except for Germany and Italy, where respondents were 75% and 82% male, respectively.

Respondents' attitude to technology is measured on a scale from -3 (technophobe) to +3 (technophile) using a questionnaire based on [17]. The mean score of our sample was 1.2 (SD 0.9, minimum -2.25, maximum 2.75), which indicates a somewhat positive attitude. This is to be expected from a younger sample responding to an online survey. Overall, 67% were concerned about their online privacy. The three main concerns were being asked for too much personal information (73%), strangers obtaining information from online activity patterns (72%), and identity theft (69%).

### Photo Preservation Practices

76% of all respondents said that preserving photos for future generations was important or very important to them, and 64% are worried or very worried about losing photos that matter.

We provided respondents with a list of photo preservation practices and asked whether these reflected their own habits. Answer options were very accurate, somewhat accurate, somewhat inaccurate, and very inaccurate. We then used cluster analysis to find groups of related response patterns. The most stable solution yielded two habit-based clusters, Group A (48%,  $n=114$ , avg. silhouette 0.17) and Group B (52%,  $n=122$ , avg. silhouette 0.21).

**Table 2:** Photo Preservation Practices, % of participants who said the statement was very or somewhat accurate. Practice descriptions are shortened for reasons of space

Practice	All	$P_A$	$P_B$
I store photos privately and securely.	67%	82%	53%
I print photos on high-quality paper.	60%	76%	45%
I keep copies in multiple places.	53%	86%	23%
I file photos carefully.	52%	74%	31%
I use automated backups.	46%	45%	47%
I keep multiple copies.	42%	75%	11%
I make manual backups.	31%	58%	7%
I check whether photos are still readable.	31%	41%	21%
I move photos to new storage media.	30%	54%	8%
I archive printed photos by scanning.	17%	28%	7%
I give copies to others for safekeeping.	15%	25%	6%

Table 2 summarises the results. While Group  $P_A$  is very diligent in transferring photos to new media, printing them off, and storing them safely, Group  $P_B$  is far more

**Table 1:** Demographics. Region is based on self-reported country of origin

Age	18–24	187	(79%)
	25–34	49	(21%)
Gender	female	160	(68%)
	male	76	(32%)
Region	UK	52	(22%)
	Sweden	74	(31%)
	Italy	22	(9%)
	Germany	16	(7%)
	Europe (Other)	27	(11%)
	Asia	12	(5%)
	Other	9	(4%)
	Not Stated	24	(10%)

sanguine. The core strategies of Group P<sub>A</sub> (*Safety in Redundancy*) all build on redundancy, guarding against the failure of a storage unit. They were far less aware of the need to regularly check whether photos are still readable. The main preservation strategies for Group P<sub>B</sub> (*File and Forget*) are secure storage, automated backups, and printing on paper.

#### *Photo Management Practices*

230 (97%) respondents take digital photos themselves. 39% download digital photos from the internet, and 34% receive them by email. Of those who take their own photos, 95% (n=219) use a smartphone, and 83% (n=190) a digital camera. 74% of smartphone users take photos on their phone daily or weekly. Digital camera owners use their cameras monthly (33%) or even more rarely (51%).

83% manage their photos using the file manager, 65% use the photo app on their smart phone, and 42% use web services such as Facebook or Flickr. The most popular computer software was iPhoto (21%), followed by Photoshop (13%).

**Table 3:** Photo Management Practices, % of participants who said the statement was very or somewhat accurate. Statements have been shortened.

Practice	All	M <sub>A</sub>	M <sub>B</sub>
I use file and folder names.	81%	75%	90%
I keep photos organised.	58%	41%	83%
I delete most digital photos.	36%	35%	37%
I label people, places, and objects.	27%	5%	59%
I add keywords and titles.	27%	3%	60%
I add information automatically.	19%	4%	40%

The question about photo management practices was

structured like the one on preservation practices. Again, cluster analysis yielded two groups, M<sub>A</sub> (n=139, 59%, avg. silhouette 0.3 ) and M<sub>B</sub> (n=97, 41%, avg. silhouette 0.22). Results are shown in Table 3. While Group M<sub>A</sub> (*Filing First*) relies almost exclusively on files and folders, Group M<sub>B</sub> (*Curators*) is more likely to curate their photos using both manual (keywords) and automatic (face recognition) options. In both groups, around a third manage the overload of digital photos through regular deletion.

## **Discussion and Implications for Design**

Our results are biased by the education level and demographics of our respondents. People who choose to fill in a survey about digital preservation are also arguably more likely to be concerned about their digital photos than people who do not. Therefore, the design conclusions we draw from our data reflects the needs of potential early adopters who are technology literate and interested in personal archiving.

### *Come to the User*

Despite this selection bias, the key message for designers is that preservation solutions should not require data to be managed by specific apps. Uptake of specialised software is low even if it comes with the operating system (e.g., iPhoto). Solutions should use file names, folder structures and the metadata automatically provided by smartphones, even though metadata quality will vary.

While almost everyone takes digital photos on smartphones, which is in line with recent statistics [11], many photos come from other sources, such as social media or photo sharing sites, that may have stripped some metadata from the original. Therefore, robust detection of groups of related photos is important.

Google Stories [8] is an interesting starting point for design. It takes photos that were uploaded to Google Plus and weaves them into a scrapbook-style story, contextualising them with date and location. Such stories could be preservation units that can be compressed further later on by paring down photos or reducing storage quality.

### Towards Personas

The average silhouette widths of each cluster indicate that there is substantial variation in our data that is not covered by the two-cluster solutions. Nevertheless, the patterns that emerge are meaningful and interpretable and can be used to develop three groups that can be the seed of personas, *Safe Curators* (n=61, 26% of our sample), *Safe Filers* (n=53, 22%), and *File and Forget* (n=86, 36%) (The fourth combination, curators who file and forget, is relatively rare (n=36, 15%)). Their views on our two key preservation questions are summarised in Table 4.

**Table 4:** Attitudes to Preservation. Loss: Worried about losing important photos; Generations: Importance of preserving important photos for future generations

Persona	Loss	Generations
Safe Curators	74%	89%
Safe Filers	68%	77%
File and Forget	56%	63%

Safety-conscious filers appear to be an ideal first target group for supported personal preservation. They do not tend to add information to their photos that supports contextualised remembering, but still care about future proofing their photo collections. The File and Forget group tends to be more sanguine about their digital photographic heritage.

### Conclusions

Data collection for the survey is estimated to finish by February 2015. Once the complete data set is available (estimated final sample size: 1,300), we will replicate our analyses to establish whether the initial personas identified hold across age groups and cultures, and examine whether particular demographic characteristics correlate with particular personas.

The next step is to conduct semi-structured interviews with a purposive sample of survey respondents who have indicated that they are willing to take part in further studies. These data will illuminate the patterns that emerge from the survey response data and allow us to flesh out the personas in greater detail.

Helping people choose what to preserve is another important avenue of future work. While Whittaker et al. [18] argue that people don't want to delete their data, a third of our sample, regardless of group, said that they deleted most of the photos they took. In earlier work, we have shown that these deletion decisions are often highly subjective [19], and need to be supported through careful interface design. The next step is to extend our work on deletion to different levels of preservation, building on the concept of preservation value that has been developed within ForgetIT [5].

### Acknowledgements

This study was funded by EU FP7 grant number 600826 (ForgetIT). We thank our collaborators in ForgetIT for their comments on earlier versions of the survey.

### References

- [1] Ames, M., Eckles, D., Naaman, M., Spasojevic, M., and House, N. Requirements for mobile photoware. *Personal and Ubiquitous Computing* 14, 2 (2009), 95–109.
- [2] Brock, G., Pihur, V., Datta, S., and Datta, S. clValid: An R package for cluster validation. *Journal of Statistical Software* 25, 4 (2008), 1–22.
- [3] Buchanan, T., Paine, C., Joinson, A. N., and Reips, U.-D. Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information*

- Science and Technology* 58, 2 (Jan. 2007), 157–165.
- [4] Kalnikaite, V., and Whittaker, S. A saunter down memory lane: Digital reflection on personal mementos. *International Journal of Human-Computer Studies* 69, 5 (2011), 298–310.
- [5] Kanhabua, N., Niederée, C., and Siberski, W. Towards concise preservation by managed forgetting: Research issues and case study. In *Proceedings of the 10th International Conference on Preservation of Digital Objects, iPres* (2013).
- [6] Logie, R. H., Niven, E., Wolters, M. K., and Mayer-Schönberger, V. Foundations of forgetting and remembering preliminary report. Tech. rep., ForgetIT Public Deliverable, 2013.
- [7] Lucero, A., Boberg, M., and Uusitalo, S. Image space. In *Proc. ACE '09* (2009), 215.
- [8] Madrigal, A. How to teach google what a story is. *The Atlantic* (2014).
- [9] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. *cluster: Cluster Analysis Basics and Extensions*, 2014. R package version 1.15.2.
- [10] Marshall, C. C. Rethinking personal digital archiving, part 1: Four challenges from the field. *D-Lib Magazine* 14, 3 (2008), 2.
- [11] OFCOM. The communications market report. Tech. rep., Ofcom, 2014.
- [12] Olsson, T., Soronen, H., and Väänänen-Vainio-Mattila, K. User needs and design guidelines for mobile services for sharing digital life memories. In *Proc. MobileHCI '08* (2008), 273.
- [13] Petrelli, D., Bowen, S., and Whittaker, S. Photo Mementos: Designing Digital Media to Represent Ourselves At Home. *International Journal of Human-Computer Studies* 72, 3 (2014), 320–336.
- [14] Petrelli, D., and Whittaker, S. Family memories in the home: contrasting physical and digital mementos. *Personal and Ubiquitous Computing* 14, 2 (2010), 153–169.
- [15] Pinheiro, E., Weber, W.-D., and Barroso, L. A. Failure trends in a large disk drive population. In *Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST07)* (2007).
- [16] Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20 (1987), 53–65.
- [17] Schulenberg, S. E., and Melton, A. M. A. The Computer Aversion, Attitudes, and Familiarity Index (CAAFI): A validity study. *Computers in Human Behavior* 24, 6 (2008), 2620–2638.
- [18] Whittaker, S., Kalnikait, V., Petrelli, D., Villar, N., Bergman, O., and Clough, P. Human Computer Interaction Socio-Technical Lifelogging : Deriving Design Principles for a Future Proof Digital Past. *Human Computer Interaction* 27, 1-2 (2012), 37–62.
- [19] Wolters, M. K., Niven, E., and Logie, R. H. The art of deleting snapshots. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (2014), 2521–2526.