Considerations on Privacy in the Era of Digitally Logged Lives

Rashmi Gupta[a], Martin Crane[b], Cathal Gurrin[a]

[a] *Insight Centre for Data Analytics, Dublin City University*
[b] *ADAPT Centre, Dublin City University*

Abstract

Purpose: The continuous advancements in wearable sensing technologies enable the easy collection and publishing of visual lifelog data. Widespread adaptation of visual lifelog technologies would have the potential to pose challenges for ensuring personal privacy of subjects and bystanders in lifelog data. This paper presents preliminary findings from a study of lifeloggers with the aim of better understanding their concerns regarding privacy in lifelog data.

Design/Methodology/Approach: In this study, we have collected a visual dataset of 64,837 images from 25 lifelogging participants over a period of two days each and we conducted an interactive session (face to face conversation) with each participant in order to capture their concerns when sharing the lifelog data across three specified categories (i.e *Private (Only for Me)*, *Semi-Private (Family/Friends)*, and *Public*).

Findings: In general we found that participants tend to err on the side of conservative privacy settings and that there is a noticeable difference in what different participants are willing to share. In summary, we found that the categories of images that the participants wished to kept private include personally identifiable information and professional information; categories of images that could be shared with family/friends include family moments or content related to daily routine lifestyle; and other visual lifelog data could potentially be made *public*).

Originality/Value: We analyse the potential differences in the willingness of 25 participants to share data. In addition, reasons for being a volunteer to collect lifelog data and how the lifelogging device affected the lifestyle of the lifelogger are analysed. Based on the findings of this study, we propose a set of challenges for the anonymisation of lifelog data that should be solved when supporting lifelog data sharing.

Paper Type: Research paper with a user study.

Keywords: Lifelogging, privacy, data anonymisation, privacy access levels

Email addresses: rashmi.gupta3@mail.dcu.ie (Rashmi Gupta), martin.crane@dcu.ie (Martin Crane), cathal.gurrin@dcu.ie (Cathal Gurrin)

1. Introduction

Lifelogging is concerned with the rich sensing of personal life experience into a digital archive [1]. It is a relatively new phenomenon in which an individual (the *lifelogger*) can track and record all their daily life activities (such as dietary routines, sleeping habits, exercise routines, social interactions, and so on), by using one or more lifelogging devices. Many such types of devices can be used for lifelogging, but perhaps the most well known is the wearable camera, as exemplified by the Microsoft *Sensecam* [2], which was used in the *MyLifeBits* project [3]. Such wearable cameras visually capture the field of view of the wearer and provide a detailed record (*log*) of the life of the individual, capturing many thousands of images autonomously per day. Other wearable devices (e.g. smartwatches) can capture additional data about the individual's activities and are aimed at interested consumers, such as the quantified self community [4]. In the case of visual lifelogging, the automated nature of the data capture and the associated lack of its manual triggering raises a number of new concerns with regard to preserving the privacy of individuals inadvertently captured in lifelog data [1]. The visual data from lifelog cameras can capture personal details of the lifelogger and also identifiable people around the lifelogger, in many cases without their consent. As a result, it may be necessary to hide the identity of captured individuals and any sensitive personal, professional, or social information before sharing this data, publishing the data, or making it available to third party organisations.

In this paper, we present a study into the privacy concerns of participants who gathered and shared visual lifelog data using off-the-shelf wearable cameras. This study has been approved by Dublin City University ethics community and collected wearable camera images from 25 lifelogging participants over a period of two days each. For the study, each participant was asked to categorise their data into three categories: data that the participant is happy to keep private (not shared with anyone); data that can be kept semi-private as such accessible by family and friends; and data that the participant is willing to share publicly (accessible by everyone). The main contributions of this paper are thus: (i) a motivation for lifelog data anonymisation to gain an understanding of the privacy concerns of individuals who gather (visual) lifelog data; (ii) an understanding of the different considerations facing a lifelogger when they want to share data such as the visual content lifelogger wish to keep private, can share with family members and friends, or can share with public; (iii) an informed list of suggested requirements for anonymisation of lifelog data; and (iv) a review of the personal experiences of the participants while collecting lifelog data.

2. History and Background

2.1. Lifelog Data

Personal data gathering for lifelogs has a long history, tracing back to Richard Buckminster Fuller's pre-digital Dymaxion Chronofile [5], which he described as a complete record of an individual, containing a chronological arrangement of

all his personal and business information, comprising thousands of papers, thousands of hours of audio and video, hundreds of models and artefacts, 1,400 feet of content and seventeen hundred hours of recordings. This physical lifelog is public, being held at the Stanford library[1]. By the early 2000s, it had become possible to capture many aspects of life experience digitally, and consequently Bell and Gemmel undertook the *MyLifeBits* project [6] to store digitally Bell's lifetime's worth of articles, books, CDs, letters, memos, papers, photographs (including periodic phases of SenseCam automatic visual lifelogging), pictures, presentations, home movies, videotaped lectures, and voice recordings. More recently, cost reduction has lead to the increased availability of wearable sensors for the lifelogger (shown in Figure 1) such as smartwatches (to monitor the biometrics of the individual), or wearable cameras (which is based on the Microsoft Sensecam) which can capture about 2,000 images every day from the viewpoint of the wearer. Additionally, smartphones carry a range of sensors, including microphones, cameras, and accelerometers, etc. Once such lifelog data is captured, it can be stored for lifelong access, or (if the lifelogger wishes) shared through various social media channels.

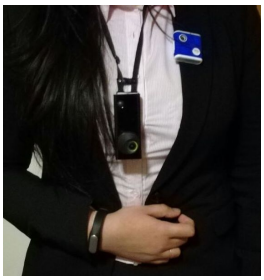

Figure 1: A Lifelogger showing a selection of wearable sensors: smartwatch(on wrist), Autographer camera(wearing around her neck), ion/snapcam camera (attached to the pocket).

Heretofore, the research community has considered lifelogging as solipsistic activity, in terms of the individual, the lifelogger. Over the past decade, one can see the increasing use of shared lifelog data as a source of evidence for epidemiological studies, notwithstanding the increasing concerns regarding personal data privacy from individuals, organisations, and societies. A review of related literature highlights the application of lifelogging tools as a means of human memory understanding [7, 8], for supporting human memory [9, 10], for facilitating large-scale epidemiological studies in health-care [11], lifestyle monitoring [12, 13], diet/obesity analytics [14], or behaviour analysis [15, 16, 17]. A typical feature of such activities is the use of wearable camera data, numbering thousands of images per day, captured by individuals in real-world settings. Therefore, one of the key motivations for this work is to gain an understanding of the privacy concerns of individuals who either gather (visual) lifelog data

[1] https://library.stanford.edu/spc/manuscripts-division/r-buckminster-fuller-timeline

for their own solipsistic purposes, or as a willing participant in organised user studies in which their data is seen by third parties or shared with third parties. We begin by exploring the concept of privacy.

2.2. Privacy and Lifelogs

Privacy, as a concept was first defined by Warren and Brandeis in 1890 as the "right to be let alone" and became of concern when photography was becoming popular as a first generation of 'portable' cameras came to market. A more contemporary definition of the right to privacy in a digital world, from [18] is: *"the right to privacy is our right to keep a domain around us, which includes all those things that are part of us, such as our body, home, thoughts, feelings, secrets and identity. The right to privacy enables us to choose which parts in this domain can be accessed by others, and control the extent, manner and timing of the use of those parts we choose to disclose"*. A key feature here is that the individual has control of his/her own domain.

The concept of privacy and data starting to gain attention from 1960's onwards; Westin [19], introduced different privacy dimensions that provide four different modes of privacy such as *Solitude*, *Intimacy*, *Anonymity*, and *Reserve*. Subsequently, Pedersen [20], conducted a user-study and identified two new additional dimensions of privacy; *Reserve* (where person shows unwillingness to interact with unknown people or strangers); *Isolation* (where person wants to be alone and away from others); *Solitude* (where person wants to be alone by oneself but free from observations by others); *Intimacy with family* (where person is interacting with their family members); *Intimacy with friends* (where the person is interacting with friends); and *Anonymity* (where the person wants to hide his/her identity). Ackerman et al. [21], discussed privacy issues while collecting human-computer interaction data. Based on individual user differences in sharing personal data, they found different types of privacy concerns including risk of unauthorised access by third-parties and the risk of reusing personal data for unrelated purposes without the consent of the data owner.

Privacy of personal data is likely to be a key concern for a lifelogger, especially so due to the always-on and passive nature of lifelog data capture. This passivity of data capture can easily lead to the accidental capture of potentially private data concerning the lifelogger themselves, or others with whom they interact. O'Hara et al. [22], proposed an idea of sharing lifelog data by integration or cross-reference with the lifeloggers themselves. Therefore, they introduced two new scopes of logging life experiences i.e. *private scope* based on storing lifelog data in personal knowledge bases and *public scope* where the lifelogger wish to share lifelog data publicly and are stored separately. Gurrin et al. [23], identified how privacy issues differ between the lifelogger and individuals captured in the lifelog and proposed a technical solution to address this challenge. Later, Hoyle et al. [24], suggested that personal data can be either private (i.e. the user does not want to share the images with family/friends/public), semi-private (either shared with selected groups such as close friends and family, other friends, or colleagues/classmates), or public (shared with anyone). Chowdhury et al. in [25], proposed a user-study with postgraduate students to understand

the potential aspects of sharing lifelog data with their online social circle and found various factors that affect the decision of sharing lifelog data such as type of content available in images (e.g. known people, identifiable objects etc.); type of activities they are performing when capturing the data; context of the images (e.g. location, sensitivity of the event in image); and the audience who can view it.

In the lifelogging domain, one may wish to remove the identity of any recognisable individuals or private/professional information of the individual from shared lifelogs, in case of harm being caused to the lifelogger or bystanders due to the sharing of data. In addition to the blurring recognisable individuals, other data may be of concern to lifeloggers, such as personally identifiable content (e.g. social security number, credit/debit card details, passport information), or other content that can identify individuals in the data (e.g. car plate number, personal messages on phones, social media shares). For examples of potentially private data and situations, see four examples in Figure 2 from [23].

In addition to the manual and survey based privacy assessment (discussed earlier in this section), Ye et al. [26], introduced the concept of negative face blurring by implementing automatic face detection, recognition and blurring to hide the unknown persons in visual lifelog data collected by Google Glass. Korayem et al. [27], proposed a new experimental approach to identify privacy concerns from readable screen content in lifelog images by deploying computer vision algorithms to automatically detect computer screens in visual lifelogs. Detection of screens in images is, of course, just part of the wider subject of automatic detection of objects in visual (non-lifelog) data, which has been the subject of much research in the computer vision field, with recent approaches typically based on the application of AI and deep learning. Lifelog data often increases the challenge for researchers due to the fact that lifelog images tend to be blurry, out-of-focus, noisy, and with significant occlusion issues, as described in [28].

Figure 2: Examples of potentially private data (A & B), Bystanders and Subjects (C & D) in the Google Glass captured lifelog data from [23]. We have explicit consent to reproduce these images from all parties (authors and individuals).

In this paper, we build on previous work and describe an interactive session with lifelogging participants to understand their concerns about sharing their lifelog data. We extend the previous work in terms of the depth and scope of the inquiry and we consider the challenge of privacy-aware lifelogging from the viewpoint of both the lifelogger and other individuals (bystanders) who are captured in the lifelog. Examples of potentially private data for the lifelogger themselves is shown in Figure 2 (A & B), which shows the reading of personal content and self-capture of the lifelogger in mirror. We also consider the privacy risks to other individuals present in the visual lifelog content such as subjects (the person who is directly in contact with lifelogger) and bystanders (the person or group of persons who are not in contact with lifelogger but available in lifelog data, potentially without their permission) shown in Figure 2 (C & D) [23].

3. User Experiment

In order to explore the issues and concerns of lifeloggers when sharing lifelog data, we conducted a user study, which was composed of three distinct stages. At the initial stage, 25 participants from different backgrounds were asked to wear an Autographer (a passive capture wearable camera) for a period of 2 days each and then asked to answer a questionnaire to verify the reasons for volunteering to participate in this user-study. In the second stage, we conducted a privacy assessment on the types of visual lifelog data that participants are willing to share, by asking them to segment their data into three categories with different access levels (or sharing levels) from private to public. The three different access levels (i.e. the categories) are:

- ***Private (Only for Me):*** The images are not to be shared with anyone, but are to be kept in a private lifelog archive just accessible by the wearer. We consider this to be private data.
- ***Semi-Private (Family and Friends):*** The images that would only be shared with family and friends, as an example of trustworthy individuals who are likely to be known to the lifelogger.
- ***Public:*** Images that can be shared publicly, such as via social networks or to researchers through some data-release process. It is important to note that this data, once shared, would likely be out of the control of the lifelogger once it is made available publicly.

The main idea behind considering each image in one of these three categories/access levels is that such a categorisation should be easily understandable, given that it is the typical privacy permissions of data in modern computing systems (i.e file or folder is being shared with owner, group, or everyone), discussed in [29] and [30]. However, it is worth noting that some data may not be suitable even for private storage, which could be dependent on the sensitivities of the lifelogger or other external/legal factors. For example, images captured accidentally in a bathroom or other private situations are likely not to

be suitable for permanent storage, even if private to the lifelogger. Such data is not considered in this research because it is deleted by the participants in a pre-experiment data cleaning phase, as described later. In the third and final stage, after completing the user experiment, each lifelogger was asked to share his/her personal experience as to how the wearable device affected their lifestyle in order to capture some guidelines regarding the development of consent based privacy-aware user friendly lifelogging frameworks.

3.1. Experimental Configuration

In order to understand the concerns of lifeloggers when sharing lifelog data, we asked each participant to donate two full days of visual lifelog data (typically between 1250 - 1750 images/day) to the main author. When carrying out this research, we adhered to ethical research policies and followed the process of lifelog data gathering in [31]. After gathering the lifelog data, all participants were allowed to remove any lifelog data, which they felt could be potentially embarrassing and offensive before engaging in the experiment.

They were then asked to review each image and make a judgement as to which categories the image should belong to, from the three categories (*P*rivate only for me i.e. full personal access control over personal data; Semi-Private *family/friends* i.e some or partial control over personal data; and *public* i.e. no control over who can access the data). Based on these findings, we would then be in a position to define a set of requirements to assist lifeloggers in choosing data to share and assist developers when choosing what anonymisation or content analysis tools to develop.

3.2. Lifeloggers/Participants

In this experiment, we gathered data from 25 participants from different backgrounds, so that we did not end up with a cohort of a single type of individual (e.g. university students). The participants were from different occupations (i.e. researchers: 10, professionals: 7, others such as undergraduates or home-makers: 8). There were 10 female participants and 15 males, all within the age range from 20-40. The three broad types of participant (researcher, professional, or other) are defined thus:

- **Researcher:** Researcher participants belong to academic environment or the general professional research community. We observe that the researchers collected a comparatively large amount of personal data by wearing the camera all-day with few interruptions. Therefore, this group tended to have a higher proportion of data labelled as private (average 40%, see Table 1).

- **Professional:** Professional participants belong to non-academic industries or organisations. They collected slightly less lifelog data than members of either of the other two categories, and they are happy to share most of their lifelog data with family/friends (average 47%, see Table 1).

Lifelogger/User	Profession	Images/2 Days	Duration/2 Days	Access Level Categories		
				Private (%)	Semi-Private (%)	Public (%)
1	Researcher	2180	28 hours	3	76	21
2	Researcher	3584	24 hours	75	24	1
3	Researcher	2156	32 hours	67	31	2
4	Researcher	2128	22 hours	5	77	18
5	Researcher	3420	26 hours	54	24	22
6	Researcher	2780	22 hours	48	28	24
7	Researcher	2992	23 hours	43	25	32
8	Researcher	2598	24 hours	55	32	12
9	Researcher	2728	27 hours	5	52	43
10	Researcher	2180	19 hours	5	51	34
Average (standard deviation)		**2,674**	**25 hours**	**40 (28.54)**	39 (20.85)	21 (13.47)
11	Professional	2610	21 hours	7	83	10
12	Professional	2175	23 hours	27	36	37
13	Professional	2523	24 hours	39	53	8
14	Professional	2863	22 hours	33	4	63
15	Professional	3264	27 hours	21	48	31
16	Professional	2384	21 hours	19	41	40
17	Professional	1925	19 hours	14	68	18
Average (standard deviation)		**2,534**	**22 hours**	23 (11.02)	**47 (25.09)**	30 (19.43)
18	Other	2491	21 hours	14	58	28
19	Other	2228	23 hours	8	34	58
20	Other	2208	24 hours	31	3	66
21	Other	1805	19 hours	27	41	32
22	Other	2992	25 hours	5	55	40
23	Other	3308	28 hours	24	52	24
24	Other	2747	21 hours	23	31	46
25	Other	2568	26 hours	9	6	85
Average (standard deviation)		**2,543**	**23 hours**	18 (09.82)	36 (21.16)	**46 (21.01)**
Overall Results (in total)		**64,837**	**591 hours**	**28**	**40**	**32**

Table 1: Summary information about participants (where participants (n)=25), lifelog dataset, and the variance of privacy preference across participants.

- **Other:** Other participants such as undergraduate university students or home-makers collect comparatively more data than professionals, but less than the researchers. Generally they collected less private content in their lifelog data and happy to publicly share most of their data (average 46%, see Table 1).

3.3. Process and Dataset

We collected 64,837 images during the experiment (25 participants for two days each) for about 12-14 hours/day. The Autographer wearable camera was configured to automatically capture 2-3 images per minute from the field of view of the lifelogger. Wearable cameras such as the Autographer have been used in many lifelog studies to identify daily life activities of the individuals, for example in privacy studies [25] or public-health studies [32] and [11]. As stated above, the participants were allowed to delete any images that they wanted from their data before donating them for this experiment. During the experiment, each participant was asked to review carefully each image and categorise it into one of three specified categories (*Private (only for me)*, *Semi-Private (family/friends)*, and *Public*). While this process was going on, the author was observing the actions of the participant who engaged with the author in a think-aloud process, explaining his/her motivations while they made decisions about each image.

4. Results and Findings

In Table 1, we show the variance of lifelog data across all three user types based on access level categories. For each user and user type, we can see the

willingness to either share or keep lifelog data private. Therefore, we observed that the researchers are willing to keep their most of the data private, professionals are happy to share their data with their family and friends and the third type (others) are happy to share much of their data to the public. In this section, we engaged in a deeper analysis of the findings presented in Table 1 and we discuss the privacy assessment by participants for their own visual lifelog data in detail, across all three experimental stages.

4.1. Stage 1: Reasons for participation in user study

In order to understand the reasons why participants had taken part in the study and their thoughts on visual lifelogging, each participant was asked the reason(s) why they voluntarily participated in this experiment and we noted a diversity in the responses, as summarised in Table 2. 76% of the participants (9 researchers, 4 professionals and 6 others) agreed to participate because they found this study to represent a novel and interesting activity, while 60% of the participants (10 researchers and 5 other participants) were prepared to collect lifelog data for self-observance and self-improvement of daily life activities. Additionally, we found that 52% of participants (3 researchers, 6 professionals and 4 others) took part in order to capture memorable experiences for future access. Perhaps due to the nature of our study which sought data during a short-time period, we observed that the participants were not concerned to collect lifelog data from any specific occasion or event, which differs from previous findings [25].

Reasons for the Participation	**Number of participants (N = 25) and Overall (%)**
New and interesting thing to do	19 (76%)
Self observance and self improvement in daily routine	15 (60%)
To recollect the memories for future	13 (52%)
Capturing/sharing important moments with your family	6 (24%)
For being social and sharing personal activities with society	4 (16%)
Just for fun	2 (8%)
May helpful for future research	1 (4%)

Table 2: Stage 1: Summary results of reasons for being participant in user study.

4.2. Stage 2: Three different privacy levels

We discuss the privacy assessment by participants for the different types of content captured in their lifelog data, by exploring their willingness to share across all the three access levels i.e. *Private (Only for me)*; *Semi-Private (Family/ friends)*; and *Public*. We do this by examining the volume of, and types of lifelog data that participants allocated to each category.

4.2.1. Private:

Table 3 provides, in decreasing order of importance, the categories of image that the participants wished to keep private.

Privacy assessment from number of participants (N = 25) for various types of content in visual lifelog data			
Content in lifelog data	**No. of participants (Private (%))**	**No. of participants (Semi-Private (%))**	**No. of participants (Public (%))**
Credit/debit card	25 (100%)	0	0
Private/professional emails on laptop/computer	25 (100%)	0	0
Cash money or ATM card usage	25 (100%)	0	0
Bathrooms	23 (92%)	2 (8%)	0
Personal communications on computer/mobile	22 (88%)	2 (8%)	1 (4%)
Images captured in a bedroom environment	20 (80%)	4 (16%)	1 (4%)
Self reflection in mirror	19 (76%)	4 (16%)	2 (8%)
Unflattering body positions	18 (72%)	5 (20%)	2 (8%)
Social media information	14 (56%)	10 (40%)	1 (4%)
Car speed/ driving style	11 (44%)	6 (24%)	8 (32%)
Paper/personal notes in physical medium	11 (44%)	10 (40%)	4 (16%)
Personal bills	10 (40%)	13 (52%)	2 (8%)
Car license plate number	7 (28%)	11 (44%)	7 (28%)
Personal physique, room or lifestyle	6 (24%)	15 (60%)	4 (16%)
Unknown people (Bystanders)	5 (20%)	4 (16%)	16 (64%)
Known person (Subject)	4 (16%)	18 (76%)	3 (12%)
Third person's house in images	4 (16%)	2 (8%)	19 (76%)
Clean home environment	3 (12%)	2 (8%)	20 (80%)
Pray	3 (12%)	2 (8%)	20 (80%)
Gardening at home	2 (8%)	2 (8%)	21 (84%)
General web surfing	5 (20%)	12 (48%)	8 (32%)
Any physical activity (exercise	4 (16%)	8 (32%)	13 (52%)
Taking rest in living room	4 (16%)	15 (60%)	6 (24%)
Food eating	4 (16%)	7 (28%)	14 (56%)
In outside environment with friends	2 (8%)	13 (52%)	10 (40%)
Normal driving images	1 (4%)	4 (16%)	20 (80%)
Shopping in grocery store	1 (4%)	3 (12%)	21 (84%)
Talking with family members	1 (4%)	20 (80%)	4 (16%)
Cooking	1 (4%)	12 (48%)	12 (48%)
Social interaction	1 (4%)	13 (52%)	11 (44%)

Table 3: Summary of the privacy assessment from the 25 participants of various types of content in visual lifelog data across three access levels: *Private*, *Semi-Private*, and *Public*, sorted in decreasing order of privacy concern.

- ***Personal and professional information:*** Unsurprisingly, all participants wished to hide their private financial details such as credit/debit card details and cash money or ATM card usage content if available in their lifelog data. The participants want to hide personal communication data (visible in the images) such as private emails, social media information, online website access or physical paper readings, personal messages over phone, car plate numbers, or personal bills (i.e. electricity, telephone, shopping or grocery bills). Some of our participants were working in professional roles and as researchers; these participants did not wish to reveal their professional information such as professional emails, meeting agendas, presentations, and professional conversations with colleagues.

- ***Environments (personal or private):*** Most of the participants in this experiment, wished to restrict access to data that may reflect badly on themselves, such as images taken in bathrooms, self-reflections in a mirror, unflattering body positions, or messy/untidy environments. It is worth noting that this is quite a broad category and varies substantially across participants, with some not caring about messy environments, but others caring a lot.

- ***Identity of the person:*** 16% of the participants wished to hide the identity of subjects (i.e people that they were directly interacting with, usually family members/friends/known person) and 20% of the partici-

pants wanted to hide the identity of the bystanders as well (i.e. people who are intentionally or unintentionally available in lifelog data without their consent).

- ***Dietary routines:*** Four participants did not wish to reveal their dietary routine with anyone. From their perspective, nobody would want to observe their food intake.

- ***Other private information:*** Some other types of private data that did not occur often in the lifelog data was still highlighted is private, such as car speed/driving style (e.g, hands on steering wheel), general web surfing, engaged in any physical activity, or images taken in anybody's home besides their own.

4.2.2. Semi-Private:

By referring again to Table 3, in decreasing order of importance, these are the categories of images that participants were generally happy to share with family and friends:

- ***Family moments:*** Most of the participants were ready to share images taken in the outside environment, or inside their home, with their family or friends. For example, talking with family members (80% of the participants), having social interaction (52%) with friends, or cooking in kitchen (48%) with family.

- ***Personal lifestyle:*** Some of the participants were happy to share the information about their personal physique, room, or lifestyle with their family or friends. For example, talking with any known person (76%), revealing personal physique, room or lifestyle (60%), having a rest in the living room (60%), hanging out with friends or colleagues outside of the work place (52%), exercising (32%), or doing some daily routine activities which would be well known by their family and friends (32%).

- ***Using laptops:*** Some of the participants (typically 48%) were happy to share images showing general web surfing, viewing media or TV, scrolling through social media content, watching movies etc., once the media does not contain personal or potentially embarrassing content.

- ***Other information shared with family and friends:*** Some other type of information available in lifelog images that was shared with family and friends such as personal bills (52%), car plate number (44%) and social media information (40%).

4.2.3. Public:

This category includes any other visual lifelog image data that does not fall into *Only for me* and *Family/friends* categories above. This data is indicative of what the participants were willing to share publicly, such as the visual content where user is gardening at home (84%), shopping in a grocery store (84%),

images with a clean home environment (80%), content in images where the user is driving (80%) or engaging in worship/ praying at home or in church (80%), the presence of an unknown person (i.e. bystanders) (64%) etc. Some participants did not mind sharing images with food intake (56%), walking or doing exercise in a park (52%), cooking with family members (48%) etc., as discussed in detail in Table 3.

An example of lifelog images associated with each access level are shown in Figure 3. This should be noted, however, that we have applied these labels on the basis of the nature of our participant cohort, which were primarily academics or professionals. It is possible that other participants, such as parents of young children (limited in our cohort) may have different opinions or viewpoints, though we believe that the general findings presented in this paper would remain similar.

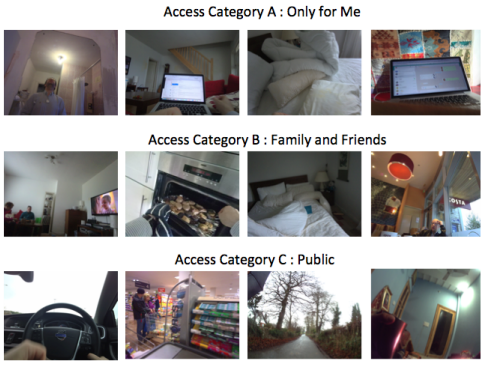

Figure 3: Example images from the specified three categories (i.e A, B and C) of anonymisation & highlight potential differences in willingness to share data i.e difference in ratio of bedroom images messiness from Access Category A: *Private* with Access Category B: *Family/ friends*

By analysing Table 3, it is possible to draw conclusions about the types of images that pose low, moderate, or high privacy concerns for all (or certain types) of participants. For example, we observe that the images showing financial information, or images taken in bathrooms carried a high degree of privacy concern for participants. At the other end of the scale, images showing driving, shopping, or social interactions posted little or no privacy concern, even if shared publicly.

Of course, this study also suggests an inherent difference between what individuals consider private. For example, some participants can share their dietary routine and their personal lifestyle publicly, but some participants wished to kept it private. Also, it is clear that there is a variance in the tolerance of different levels of messiness in the home between participants. Even within participants, we can see a difference in tolerance. For example, two different images of the same bedroom in which one image with messy content is kept private while the other image with clean bedroom (as per user's viewpoint) is shared with family/friends, shown in Figure 3. Hence, it is important that individuals who

gather such data are supported in easily defining what is to be shared and what can be kept private, or semi-private. In any case, it would be appropriate to err on the side of caution and assume a strong privacy requirement.

4.3. Stage 3: Experiences of the Wearers while Lifelogging

Each participant was then asked about their personal experience while collecting lifelog data 10 days after completion of the experiment. It was our belief that this short separation between data gathering and interview session would facilitate a better understanding of the overall experience of the lifelogging. Findings suggests that more than half (56%) of the participants experienced an enhanced realisation about their daily lifestyle such as food habits, exercise routine or working hours (i.e. self-realisation) from the two days of lifelogging. Additionally almost half of the participants (48%) felt self-motivated and happy to review their lifestyle and activities (i.e. self-reminiscence). A similar number of participants reported a self-motivated pressure to remember turn off the camera in bathrooms or while engaged in any intimate moments (i.e. self-control). Some participants (20%) felt that the lifelogging device affected their activities (i.e. influencing activities), for example the participants started to go for exercise in the gym regularly and started having healthy food in an organised manner. Six of the participants (all with professional background) were concerned to see so much confidential information such as ATMs or personal notes when reviewing their data (i.e. self-agitation). Five of the participants who belong to research community reported that data gathering was a normal part of their daily life and they did not even notice wearing camera. While one participant commented that he/she did not notice the device while busy in work, but once outside of work, there was an increased awareness of the device and an experience of difficulty in relaxing (i.e. discomfort) in social environments. However, most of the participants found that this experiment could provide insights for self-observance and self-knowledge in daily routine while some expressed surprise that they found it very useful for the self-observance in their daily life activities and behaviour.

4.4. Examining the Differences between Participants

In this preliminary study we found that (on average) 72% of collected lifelog data could potentially be shared to associated group of persons (i.e. 40% for Semi-Private category) or to third party organisations (i.e. 32% for Public category), see Figure 4.

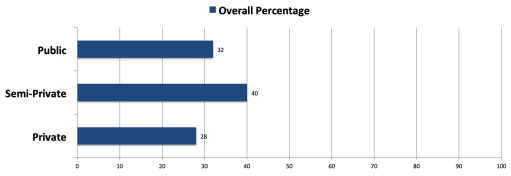

Figure 4: Overall percentage of lifelog data shared with family/friends and public.

In Figure 5, we present each judgement category on a per-participant basis, in decreasing order of privacy expectation. We observe the first 6 participants (i.e. participants 2, 3, 8, 5, 6, and 7 from Table 1) who belong to research community, wanted to make their lifelog data relatively more private. It is our conjecture that this is because academic researchers are likely to have a higher-degree of awareness concerning issues of privacy and personal data. The participants (1 and 9), both of who are researchers but were at the very early stages of their careers supported this conjecture. We can see a difference between the most privacy-aware participant (user 2 who keeps 75% of lifelog data as private) and the least concerned participant (user 1 at 3%).

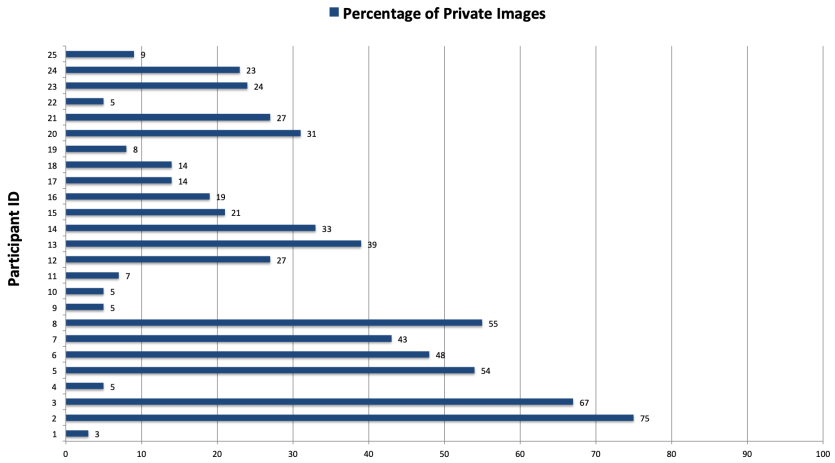

Figure 5: Percentage of lifelog data the participants wished to keep private.

Next, considering the number of images in the semi-private category in Figure 6, we observe 4 participants (i.e. participants 11, 4, 1, and 17) who belong to research and professional backgrounds were happy to share a large percentage of their visual lifelog data (images taken in home environment, social media information, self-reflection in mirror or personal physique, room or lifestyle) with family and friends (i.e. 83%, 77%, 76%, and 67%). On the other hand, three participants (i.e. participants 25, 14, and 20) who belong to other occupations (i.e. undergraduate students or home makers) were significantly less willing to share the images with family and friends (i.e. 6%, 5%, and 3%) in case where the lifelogger engaged in cooking, eating, personal/private chat on computer/mobile phone, or images taken in bathroom. The average here is 40%.

Finally, observing Figure 7 (unsurprisingly, given the previous figures), we note that 5 participants (i.e. participants 25, 20, 14, 19, and 24) with different occupations (mostly undergraduate students) were willing to share a large proportion of their data to the public domain and have a lower proportion of private content (i.e. 84%, 66%, 63%, 58%, and 46%, respectively). The participants who share less publicly and have more privacy threatening work-related content

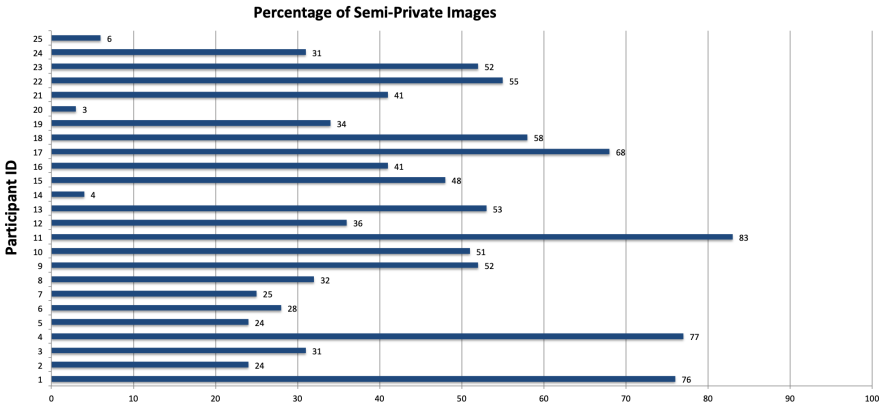

Figure 6: Percentage of lifelog data the participants wished to share with family and friends.

in their lifelog collection (i.e. the participants 8, 11, 13, 3, and 2 who share 12%, 10%, 8%, 2%, and 1% of their data respectively) are mostly researchers and professionals, with an average public data percentage of only 32%. It is our conjecture that there is a clear separation in the privacy awareness and concerns between individuals who have been born-digital and those who have experienced digital and social media content later in life.

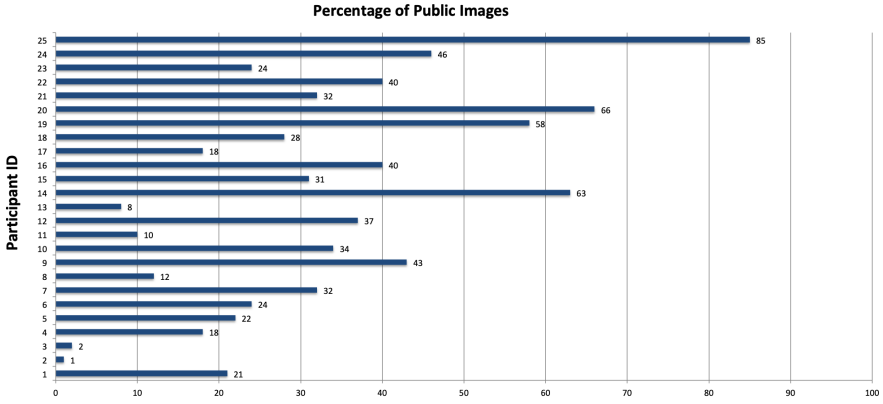

Figure 7: Percentage of lifelog data the participants wished to share with public.

In addition, we have observed (from Table 1) that the participants belonging to the research community had a higher desire to keep data private, when compared with other participant types. In summary, researchers want to keep most of their visual lifelog data private; professionals were happy to share most of their visual archive with family and friends while other participants have less private concerns and are willing to share most of their lifelog data with public or third party organisations (see Figure 8). Although this is a small preliminary study, we do note that there are differences on display that are worthy of further

study with larger cohorts.

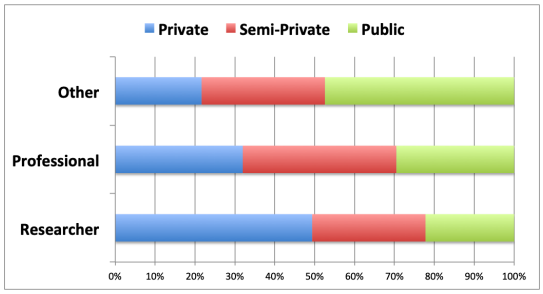

Figure 8: Overall summary of the privacy concerns across participants.

5. Discussion and Suggestions for Future Research

The experiment outlined in this paper presents a motivation for lifelog data anonymisation to gain an understanding of the privacy concerns of lifeloggers who gather lifelog data to share with family and friends, or with third parties. In common with the previous work [24, 25, 33], we investigated and analysed the sensitive content available in the visual images that stated the sharing preferences of participants (results shown in Table 3 earlier). Unlike the previous studies [24, 25], where visual lifelog data was captured from a small group of university students over a short period of time, we have captured 50 days of visual lifelog data (in total) from 25 participants with various different professions. We found significant differences between our participants in terms of what they are willing to share, or keep private, as summarised in Table 1. We found some different results from Hoyle et al. [24] and Chowdhury et al. [25], in particular, the participants (typically 48%, mostly university students and early-stage researchers) in our study were happy to share images showing general web surfing, viewing media or TV, scrolling through social media content, watching movies etc., once the media does not contain personal or potentially embarrassing content. This would differ from the more experienced professionals, who were comparatively more privacy-aware for sharing their lifelog data.

Based on this initial study, we can define a set of recommended principles for researchers using lifelog data for their epidemiological studies, or researchers who wish to release real-world lifelog collections for the community.

- We found a wide variance in privacy sensitivity between individuals. Hence, we suggest that participants gathering data for lifelog experiments should be given the opportunity to remove any sensitive data before sharing it with researchers, even if the data will be fully anonymised. This is in line with recommendations for using lifelog data in academic studies in [34].

- Due to our observations on the sensitivity differences among our cohort, we suggest that participants in lifelogging studies should be involved in

defining how/what data from their lifelog should be released to trusted parties, or released more generally in datasets.

- De-identification of the data, which is a standard pre-release processing step for visual lifelog data [31], is not a panacea when sharing data and it will not solve the privacy problem for all lifeloggers. This is because semantic judgements on user context (e.g. messy room) are key factors, and not simply the presence of faces or certain identifiable objects. Our understanding is that there are not a suite of visual concept detectors available that are trained to identify some of the more semantically nuanced issues that we have found, such as the difference between tidy and messy rooms, reflections in mirrors, activities in the bathroom, etc.

- Trusted-researcher agreements that clearly outline the expectations on researchers to respect the privacy of donating lifeloggers and implement appropriate data governance methodologies are also highly recommended. This is the process that was followed for the NTCIR lifelog datasets that were released in recent years [35].

6. Conclusion and Future Work

In this preliminary study, data was collected from a cohort of participants in an effort to better understand the sensitivities to sharing of data with friends/ family and publicly. It was found that sensitivities varied enormously across the cohort, with half willing to share more than 2/3 of the daily lifelog data publicly, while others are willing to share less than 1/4 of the data publicly. Similar differences exist in terms of sharing with friends/family, or keeping the data totally private.

Although this is a preliminary study, it allowed us to draw some initial conclusions, which would be helpful for practitioners or researchers gathering datasets for release or for user studies. We suggested a number of principles for organising and releasing lifelog data, such as the need for the lifelogger to review their content pre-release, or the lack of available accurate automated tools to automate the anonymisation process.

However, we are aware that there are many limitations of this initial experiment that need to be taken into account. The number of participants was small at 25 and there were limited professions included in our cohort, which could affect the findings. It is possible that a different cohort could present some different findings in terms of the types of content can be shared. Similarly, our analysis of inter-cohort differences in terms of profession could differ with a larger or different cohort. Another limitation is that some of the participants may already have been more diligent than others in filtering their lifelog data before starting the experiment, thereby reducing the volume of data that is likely to be labelled as private, although we do not believe that this has had a significant impact on the findings, since any such filtered data was unlikely to have been categorised for family/friends or public consumption.

There is also a subtle concern from the fact that the identity of the participant was be known to the researcher who lead this paper, so this could have impacted on the willingness of participants to share certain types of data. Had the authors been in a position to assure complete anonymity, then some of the concerns seen in this paper could potentially be lessened. However, one caveat to note here is that even if the experiment is performed anonymously, the potential for identification of the participant (or a lifelogger) through some technical advancements in the future can not be denied, so the authors suggest to err on the side of caution and to assume that the identity of a participant in a released dataset or collection is likely to be known at some point in the future, regardless of the steps taken to ensure anonymity.

A final point to raise for future work is that we only considered visual lifelog image data for this experiment; there may be other privacy concerns raised by other data sources, such as location, heart rate, URLs visited, etc. Future work should consider this issue also. It is our plan to enhance this preliminary study with a larger experiment with a more diverse cohort, so that we can validate these initial findings and engage in a more in-depth analysis.

7. Acknowledgement

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289.

References

[1] C. Gurrin, A. F. Smeaton, A. R. Doherty, Lifelogging: Personal big data, Foundations and Trends in Information Retrieval 8 (2014) 1–125.

[2] V. Bush, As we may think, The Atlantic Monthly 176 (1945) 101–108.

[3] G. Bell, J. Gemmell, A digital life, Scientific American 296 (2007) 58–65.

[4] J. Meyer, S. Simske, K. A. Siek, C. G. Gurrin, H. Hermens, Beyond quantified self: Data for wellbeing, in: CHI 14 Extended Abstracts on Human Factors in Computing Systems, CHI EA 14, Association for Computing Machinery, New York, NY, USA, 2014, p. 9598.

[5] R. G. T. Hsiao-yun Chu, Buckminster Fuller: A Technocrat for the Counterculture, 2009.

[6] J. Gemmell, G. Bell, R. Lueder, MyLifeBits: a personal database for everything, Communications of the ACM 49 (2006) 88–95.

[7] F. Milton, N. Muhlert, C. R. Butler, A. Smith, A. Benattayallah, A. Zeman, An fMRI study of long-term everyday memory using SenseCam., Memory 19 (2011) 733–744.

[8] P. J. Barnard, F. C. Murphy, M. T. Carthery-Goulart, C. Ramponi, L. Clare, Exploring the basis and boundary conditions of SenseCam-facilitated recollection, Memory 19 (2011) 758–767.

[9] E. Berry, N. Kapur, L. Williams, S. Hodges, P. Watson, G. Smyth, J. Srinivasan, R. Smith, B. Wilson, K. Wood, The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: a preliminary report., Neuropsychological rehabilitation 17 (2007) 582–601.

[10] M. Harvey, M. Langheinrich, G. Ward, Remembering through lifelogging: A survey of human memory augmentation, Pervasive and Mobile Computing 27 (2016) 14–26.

[11] L. N. Signal, M. B. Smith, M. Barr, J. Stanley, T. J. Chambers, J. Zhou, A. Duane, G. L. S. Jenkin, A. L. Pearson, C. Gurrin, A. F. Smeaton, J. A. Hoek, C. N. Mhurchu, Kids'cam: An objective methodology to study the world in which children live., American journal of preventive medicine 53 3 (2017) e89–e95.

[12] G. Wilson, D. Jones, P. Schofield, D. J. Martin, The use of a wearable camera to explore daily functioning of older adults living with persistent pain: Methodological reflections and recommendations, Journal of Rehabilitation and Assistive Technologies Engineering 5 (2018) 205566831876541.

[13] T. H. C. Nguyen, J. C. Nebel, F. Florez-Revuelta, Recognition of activities of daily living with egocentric vision: A review, 2016.

[14] Q. Zhou, D. Wang, C. N. Mhurchu, C. Gurrin, J. Zhou, Y. Cheng, H. Wang, The use of wearable cameras in assessing children's dietary intake and behaviours in China, Appetite 138 (2019).

[15] B. Everson, K. Mackintosh, M. Mcnarry, C. Todd, G. Stratton, Can wearable cameras be used to validate school-aged childrens lifestyle behaviours?, Children 6 (2019) 20.

[16] N. Li, C. Gurrin, M. Crane, H. J. Ruskin, NTCIR-12 lifelog data analytics, in: Proceedings of the first Workshop on Lifelogging Tools and Applications, LTA@MM 2016, Amsterdam, Netherlands, October 15 - 19, 2016, pp. 27–36.

[17] N. Li, M. Crane, C. Gurrin, H. J. Ruskin, Finding motifs in large personal lifelogs, in: Proceedings of the 7th Augmented Human International Conference, AH 2016, Geneva, Switzerland, February 25-27, 2016, pp. 9:1–9:8.

[18] Y. Onn, Y. Druchman, R. Timor, A. Maroun, Y. Nachmani, S. Sichlai, M. Fishman, M. Geva, A. Zyssman, I. Lev, T. Maron, Y. Simsolo, A. Fuches, S. Packer, Privacy in the digital environment (2005).

[19] A. F. Westin, Privacy and freedom, in: Washington and Lee Law Review, 25(1), 166.

[20] D. M. Pedersen, Dimensions of privacy, in: Perceptual and Motor Skills, 48(3c), pp. 1291–1297.

[21] M. S. Ackerman, S. D. Mainwaring, Privacy Issues and Human-Computer Interaction, O'Reilly Media, Cambridge, MA, pp. 19–26.

[22] K. O'Hara, M. M. Tuffield, N. Shadbolt, Lifelogging: Privacy and empowerment with memories for life, in: Identity in the Information Society, 1 (155), pp. 155–172.

[23] C. Gurrin, R. Albatal, H. Joho, K. Ishii, A privacy by design approach to lifelogging, in: In: O Hara, K. and Nguyen, C. and Haynes, P., (eds.) Digital Enlightenment Yearbook, 2014, pp. 49–73.

[24] R. Hoyle, R. Templeman, D. Anthony, D. Crandall, A. Kapadia, Sensitive lifelogs: A privacy analysis of photos from wearable cameras, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1645–1648.

[25] S. Chowdhury, M. S. Ferdous, J. M. Jose, Exploring lifelog sharing and privacy, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, pp. 553–558.

[26] T. Ye, B. Moynagh, R. Albatal, C. Gurrin, Negative faceblurring: A privacy-by-design approach to visual lifelogging with google glass, in: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 14, Association for Computing Machinery, New York, NY, USA, 2014.

[27] M. Korayem, R. Templeman, D. Chen, D. Crandall, A. Kapadia, Enhancing lifelogging privacy by detecting screens, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI 16, Association for Computing Machinery, New York, NY, USA, 2016, p. 43094314.

[28] C. Gurrin, A. F. Smeaton, D. Byrne, N. O'Hare, G. J. F. Jones, N. O'Connor, An examination of a large visual lifelog, in: AIRS 2008: Information Retrieval Technology, Springer Berlin Heidelberg, 2008, pp. 537–542.

[29] J. Phelps, G. Nowak, E. Ferrell, Privacy concerns and consumer willingness to provide personal information, Journal of Public Policy & Marketing, Privacy and Ethical Issues in Database/Interactive Marketing and Public Policy 19 (2000) 27–41.

[30] A. L. Cutillo, R. Molva, T. Strufe, Safebook: A privacy-preserving online social network leveraging on real-life trust, in: IEEE Communications Magazine, December, pp. 94–101.

[31] D.-T. Dang-Nguyen, L. Zhou, R. Gupta, M. Riegler, C. Gurrin, Building a disclosed lifelog dataset: Challenges, principles and processes, in: Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, CBMI '17, ACM, New York, NY, USA, 2017, pp. 22:1–22:6.

[32] M. Barr, L. Signal, G. Jenkin, M. Smith, Capturing exposures: using automated cameras to document environmental determinants of obesity, Health Promotion International 30 (2014) 56–63.

[33] B. A. Price, A. Stuart, G. Calikli, C. Mccormick, V. Mehta, L. Hutton, A. K. Bandara, M. Levine, B. Nuseibeh, Logging you, logging me: A replicable study of privacy and sharing behaviour in groups of visual lifeloggers, Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 1 (2017).

[34] P. Kelly, S. J. Marshall, H. Badland, J. Kerr, M. Oliver, A. R. Doherty, C. Foster, An ethical framework for automated, wearable cameras in health behavior research, American Journal of Preventive Medicine 44 (2013) 314 – 319.

[35] C. Gurrin, H. Joho, F. Hopfgartner, L. Zhou, R. Albatal, Ntcir lifelog: the first test collection for lifelog research, in: SIGIR'16 - the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, pp. 705–708.