A Privacy by Design Approach to Lifelogging

Cathal GURRIN [a,1], Rami ALBATAL [a], Hideo JOHO [b] and Kaori ISHII [b]
[a] *Insight Centre for Data Analytics,*
Dublin City University, Dublin, Ireland
[b] *University of Tsukuba,*
Tsukuba, Japan

Abstract. Technologies that enable us to capture and publish data with ease are likely to pose new concerns about privacy of the individual. In this article we examine the privacy implications of lifelogging, a new concept being explored by early adopters, which utilises wearable devices to generate a media rich archive of their life experience. The concept of privacy and the privacy implications of lifelogging are presented and discussed in terms of the four key actors in the lifelogging universe. An initial privacy-aware lifelogging framework, based on the key principles of privacy by design is presented and motivated.

Keywords. Lifelogging, privacy, privacy by design, negative face blurring

1. Introducing the Concept of Privacy Aware Lifelogging

Recent inventions and business methods call attention to the next step which must be taken for the protection of the person, and for securing to the individual what Judge Cooley calls the right *to be let alone* [2]. Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that *what is whispered in the closet shall be proclaimed from the house-tops* [1]. The above quote from Warren and Brandeis, writing about privacy and the law in the Harvard Law Review, was from 1890. At that time, the ready availability of portable cameras and means of publishing was considered a threat to privacy.

Consider these comments today in light of a new privacy threat that is coming from technologies that enable the individual to capture data and publish it with ease. In fact, such new technologies appear with increasing frequency, sometimes too fast for us to even comprehend their long-term social impact before they become commonplace. One current example is the wearable computing trend, in which individuals can wear computing systems designed to sense the environment of the individual for some potential benefits. Consider, for example, that a wearable pedometer using inbuilt accelerometers can

[1] Corresponding Author: Cathal Gurrin, Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland; E-mail: cathal@gmail.com.
[2] from Cooley on Torts, January 1878, P 29.

discover statistics about the ambulatory activities of the individual. Such technologies are becoming pervasive now, and the so-called Quantified Self [2] movement is developing in this marketplace.

However, a new generation of outward looking wearable device has recently matured and come to market. These devices incorporate wearable cameras (among other sensors) to capture detailed photo and video logs of a person's activities. Devices such as the Narrative Clip wearable camera, or Google Glass, inevitably raise privacy concerns as we see the emergence of sousveillance [3] as an increasingly popular activity.

This new generation of wearable devices allows the individual to record aspects of their life in previously unimaginable detail. Whether it is every heart beat (the Basis watch), our locations and motion (the Moves app), or everything we see and do (the Narrative Clip), we are at a new phase of technical advancement that provides us with new and potentially valuable information about ourselves. Should we so wish, we can now record video of our every waking moment and store it indefinitely. This concept of recording life in image, video and sensor data has already been started by some early adopters and it is called Lifelogging [4]. Lifelogging has recently become technically feasible, and early adopters such as Gordon Bell [5], Steve Mann [6], Cathal Gurrin [7] and Kioharu Aizawa [8] are exploring the potential of what is possible and uncovering the problems and challenges which need to be addressed. In fact, Gurrin's experiences directly influence this work.

These early adopters saw lifelogging as a solipsistic (for one's own purposes) activity, one in which the lifelog that is gathered is made available only to the lifelogger who can extract the potential benefit from the lifelog data. Since these lifeloggers wear cameras that can capture thousands of images or hours of video daily, the potential exists for significant privacy concerns to be raised. In this chapter, we are concerned with the privacy of the individual, but more so with the privacy of those with whom the individual interacts. We inform this discussion by personal experience of eight years of continual lifelogging, as well as integrating the thoughts and opinions of concerned parties. We discuss what we consider privacy to mean, present the motivation for privacy by design lifelogging and propose a privacy-aware lifelogging framework that provides the benefits of lifelogging, but maintains privacy policies designed to protect the privacy of subjects and bystanders in the lifelogger's data. As such, the focus of this article is on the extreme end of the lifelogging scale; the form of lifelogging that attempts to gather a media rich archive of life experience, as envisaged by the early adopters.

While, for many readers, widespread societal adoption of such an extreme form of lifelogging may not appear likely, the fact that it is even possible suggests that we should consider the implications. Consider for a moment that the ubiquity of mobile phones, or the popularity of social networks, would once have appeared outlandish, yet most people with mobile phones are probably members of a billion-user social network and may share intimate details of their lives with (on average) 338 social network contacts. In this article, we consider the privacy implications of a scenario in which individuals gather detailed lifelogs for their own access and benefit. Related issues such as lifelog sharing, post-life-log management, and technical issues related to data security are not included in this discussion.

2. An Introduction to Lifelogging

2.1. Defining lifelogging

Lifelogging represents a phenomenon whereby individuals can digitally record their own daily lives in varying amounts of detail and for a variety of purposes. In a sense it represents a comprehensive black-box of a person's life activities and offers great potential to mine or infer valuable knowledge about life activities. Typically, early adopters to lifelogging considered it to be an activity that was engaged in by the individual for their own benefit, and many would even counsel against sharing lifelog data. However, if lifelogging becomes more pervasive, one can imagine that many users would be willing to share aspects of their lifelog [9], perhaps via social relationships or online friend networks. We don't take this viewpoint in this article; rather we focus on the idea that lifelogging is for the lifelogger and we define our contribution accordingly.

Although there are many definitions in literature, we define lifelogging to be: *a form of pervasive computing which utilises software and sensors to generate a permanent, private and unified multimedia record of the totality of an individual's life experience and makes it available in a secure and pervasive manner*. This definition is based on a prior definition [10] which focuses on the data-capture only. In our revised definition, we include the idea that the process of lifelogging should consist of of data gathering, storage, analysis and access.

A key aspect of this definition is that the lifelog should archive the totality of an individuals experiences, i.e. follow Bell and Gemmels' vision of total capture [11]. This means that engaging in the process of lifelogging will result in the capture of significant quantities of rich multimedia data about the lifelogger, but by necessity, also about the environment of the lifelogger, and the people and objects contained therein. It is important to consider that lifelogging is typically carried out ambiently, or passively, without the lifelogger having to initiate anything, and it is different to the current mass data acquisition activities of some web organisations. The fact that lifelogging captures data ambiently results in the additional problems of non-curation or non-filtering, and the individual lifelogger may not even be aware of all the data - or the implications of keeping this data - that exists in the lifelog.

Lifelogging becomes possible as a result of three parallel advances in technology. Firstly sensors are becoming cheap, reliable, robust, power-efficient and portable. Our mobile phones pack enough sensors to digitally lifelog our activities in great detail [12,13]. Secondly, data transmission and storage technologies means that it is cheap to transmit, store and share this new sensor data. From [4] we know that we can currently store 6-8 years worth of wearable camera images on a $100 hard drive. Thirdly, we have new search and artificial intelligence techniques to allow us to convert large volumes of raw sensor data into meaningful semantic information that can provide the individual with new knowledge. These three technological advancements lead us to this point; the advent of the era of the lifelogged individual.

It is reasonable to assume that the current state of technical advancement will not slow down. Rather information technology is, as Ray Kurzweil (futurist and a director of engineering at Google) explains, on an exponential pathway of advancement [14]. There have been many predictions and laws declared in this regard, such as Moore's Law [15] for CPU transistor densities, Kryder's Law [16] for disk storage size, and Kurzweil's Law

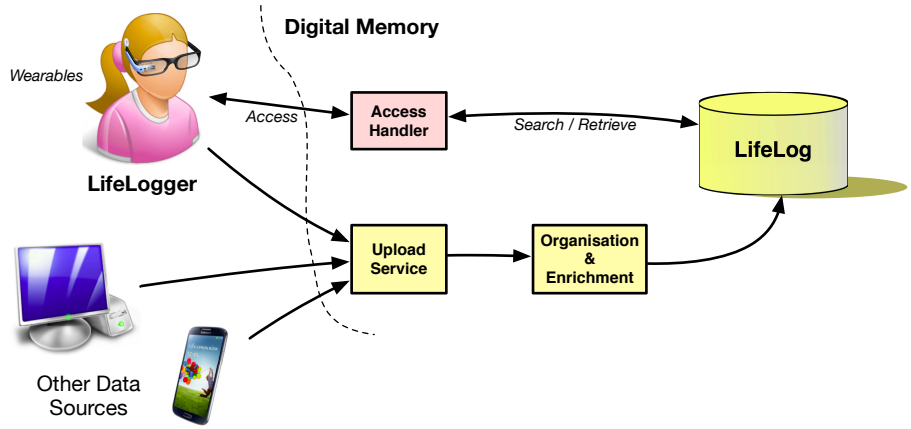

Figure 1. A basic lifelogging architecture showing the Lifelogger, the Lifelog and the Digital Memory.

of Accelerating Returns [14] for general information technology advancement. There is now real potential that these technological advancements will usher in an era of, what Gordon Bell and Jim Gemmell refer to as *total capture* lifelogging, in their book "Your Life Uploaded"[3] [11]. As the ready availability of lifelogging devices increases and people begin to see the benefits of maintaining a lifelog, then it has potential to become a common activity.

There are many potentially life enriching benefits for the individual of a lifelogging world, such as the potential for better self-awareness leading to longer and more active lifespans, new personalised healthcare applications, enhanced methods of learning, increased productivity in the workplace, increased independence or mobility for people suffering from various memory and cognitive impairments, and new forms of offline and online social interaction. As a result, society would also benefit from a healthier and more productive population.

However, this progress also raises new questions about the expectation of privacy for the lifelogger and those around them. A simple lifelog in terms of personal activity/performance monitoring using sensing techniques devoid of microphone and camera is likely to pose only minor risks for the privacy of individuals in the environment of the lifelogger [4]. However, it is our conjecture that the real challenge occurs when the lifelogger follows Bell & Gemmels vision of total capture lifelogging, which can utilise a rich range of sensors including cameras and microphones to generate a detailed trace of the lifeloggers activities.

Because lifelogging is an emergent area, it is full of terminology that is not well considered and defined. Therefore, for the purposes of this discussion, we regard the lifelogging process as having the following three core elements (see Figure 1).

[3] Your Life Uploaded was previously called Total Recall.

[4] Of course, this is a simplistic view, but one that we feel holds in most cases. Quantified Self data gathering, where the individual senses inwards would not explicitly record the presence of others, yet if multiple datasets from multiple individuals were examined simultaneously, GPS co-location, Bluetooth device interactions, etc. could lead to unforeseen privacy issues emerging.

- Lifelogging is the process of passively gathering, processing, and reflecting on life experience data collected by a variety of sensors, and is carried out by an individual; the lifelogger. This data is uploaded into the lifelog that resides in the digital memory.
- A lifelog is the actual data gathered. It could reside on a personal hard drive, in the cloud or in some portable storage device. In this discussion, we take the viewpoint that the lifelog resides in a dedicated software solution called the digital memory.
- A digital memory is akin to a digital library; it is the data from the lifelog and the associated software required to organise, semantically enrich and manage that lifelog data. It is in the digital memory that lifelog data privacy policies would need to be applied.

Since a passively gathered lifelog would contain media-rich imagery and sensor data, it is our belief that the lifelogging software (the digital memory) needs to implement privacy-aware policies to protect both the lifelogger and the people with whom that the lifelogger interacts with, henceforth called the lifelogging actors. We will now explore the nature of lifelog data and the various actors who are part of a lifelogging scenario.

2.2. Lifelogging Data

Since lifelogging is concerned with passively sensing the totality of an individual's life experience, there are likely to be a wide range of lifelogging tools employed at this point in time. As a starting point, we refer to [17] for an overview of the different categories of lifelogging tools that have been employed. We have taken the tools listing from [17] and extended it, based on our own experiences with lifelogging. These categories are described below.

- Passive Visual Capture. Utilising wearable devices such as the OMG Autographer, Narrative Clip and previously the Microsoft SenseCam [18], allows for the continous and automatic capture of life activities as a visual sequence of digital images (up to 4,000 per day [19]). These wearable cameras are typically worn on a lanyard around the neck, or clipped to clothing and as such, capture images from the viewpoint of the individual at a frequency of several each minute. For an analysis of the types of visual data gathered by these cameras, see [19]. Smartphones running dedicated software [12,13] in a suitable orientation, or wearable devices such as Google Glass, can also capture a rich visual recording of life activities from the point-of-view of the wearer. Such devices have the potential to gather upwards of 500GB of image data per year [4], which can comfortably fit onto most computer hard drives available nowadays. Images are only one aspect of visual capture and although image capture has been the focus of most lifelogging research up until now, the concept of wearable video capture is becoming a reality with many recent commercial offerings, principally aimed at sports applications. Migrating to video capture increases the richness of the multimedia archive, though it also increases the data volumes significantly. Until now, limitations in battery technology and the disk space required (32TB+ per year for HD video [4] have meant that all day video lifelogging is not considered feasible yet[5].

[5]The authors recent experiments with wearable 4K cameras (Panasonic A500) suggest a data capture rate of about 20GB per hour, or 114TB per year is required for extremely detailed video at 3840 x 2160 resolution.

It is likely that the public awareness of visual lifelogging will increase with the availability of Google Glass [20] and other tools such as the Narrative Clip. In the experience of the authors many people are not aware of wearable cameras yet and those that are, are more concerned about audio and video capture rather than a continual stream of photos.

- Passive Audio Capture. Although less explored that passive capture visuals, any smartphone can capture all-day audio without a major impact on battery life. The data requirements are about 1GB per day [4] so this will not cause any major storage problems either. Audio capture could allow for the identification of spoken words, identification of events of interest using audio event detection and potentially for the identification of who was speaking (from a limited list of friends). In [21], a continuous audio lifelog is captured, segmented and annotated with acoustic and semantic tags. [6]
- Personal Biometrics. There are many personal sensing devices for monitoring everyday activities and aimed at the consumer market and these are used by interested parties, such as the quantified self community [22]. Such devices monitor human performance, for example activity levels (number of steps taken, distance travelled, caloric output), sleep duration, etc. Since these devices only examine the individual, and don't look outwards, the dominant privacy concern is for the individual data gatherer.
- Mobile Device Context. This refers to using the smartphone to continuously and passively capture the users context, as the smartphone can now be used to record location, acceleration and movement, WiFi signal strength, and various other sensors. Coupled with a new generation of smart watches, the smartphone will be able to capture much of life activity. With power-aware sensing, it is possible to gather a contextual lifelog for an entire day without impacting too negatively on battery life. In this case, the impact on privacy is mostly going to be related to the individual lifelogger, unless scenarios occur in which multiple lifelogs may be accessed together.
- Communication Activities. Passively logging our (electronic) communications with others - such as our SMS messages, instant messages, phone calls, video calls, social network activities and email exchanges - is also possible and can form part of a lifelog. Some content, such as SMS messages, instant messages, emails and social network feeds, are naturally stored indefinitely and referred to as needed. However, phone calls and video calls are currently seen in a different light. In some jurisdictions, it is permitted to record voice communications with the agreement of one of the two parties, but in other jurisdictions, both parties must agree. This poses interesting privacy concerns for those communicating with a lifelogger who is recording communications.
- Data Creation/Access Activities. Logging our data consumed and created; all the words typed, web pages read, YouTube videos watched and so on. The Stuff-Ive-Seen system from Microsoft Research [23], is an example of such a logging tool, as is the work by dAquin et al. on monitoring and logging web information ex-

[6] In the author's experience, audio recording does pose concerns for those that the lifelogger interacts with. These bystanders are concerned that their words may be replayed in an unfavourable manner at some later point.

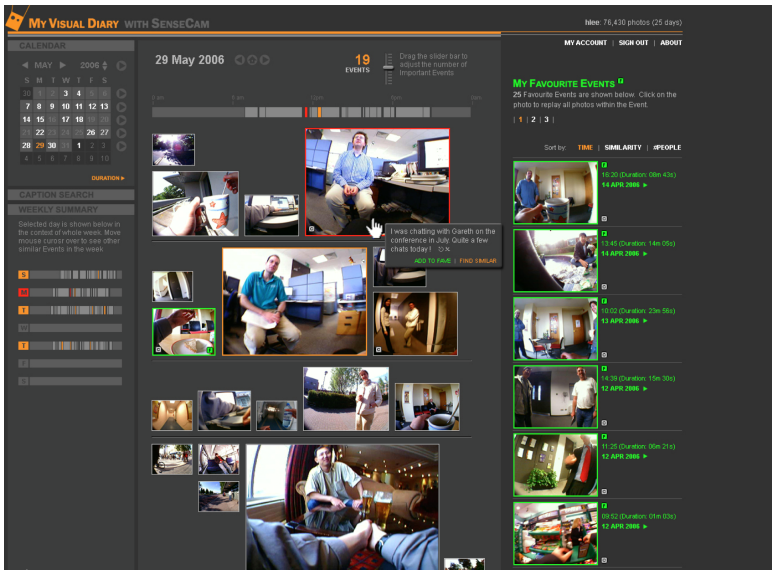

Figure 2. A typical lifelog browser from [27].

changes [24]. This area is sometimes known as Personal Information Management [25], and generally only focuses on an individuals web/PC interactions.

- Environmental Context and Media. Lifelogging is mostly, but not exclusively about recording using wearable technology. Sensors in the home, or surveillance cameras in the environment, can capture user activities and environmental context. For example, a system for retrieval and summarisation of multimedia data from a home-like environment that generates a life-log of the home [26] and the activities that occur therein.
- Manual Logging Life Activities. This refers to the indirect or direct logging of activity that is initiated by the user, and covers manual data capture, such as photo taking and recently popular services such as video clip-per-day and selfie-photo capture per day services.

While much of this lifelogging data is not going to pose major privacy concerns for either the lifelogger or the people around them, some outward-looking data, such as images and audio will naturally pose a concern. Indeed, the more comprehensive and continuous the lifelogging, the more significant the ethical and legal problems become [28]. Society is not surprised when someone keeps all their email communication or backs-up their SMS messages to the cloud before downloading them onto a new phone. However, the move to visual and aural lifelog data is a new departure and one that is likely to cause considerable unease, at least until positive benefits of lifelogging could make it an activity that the majority are willing to engage in.

Let us consider the data gathered and presented in one of the earliest and best known lifelog browsers, developed in 2006, and shown in Figure 2. Here the lifelogger is presented with software that summarises thieir day (3,000+ images) into a sequence of events and makes them accessible via a web interface. The relative importance of each event is determined algorithmically and selecting any event presents the lifelogger with

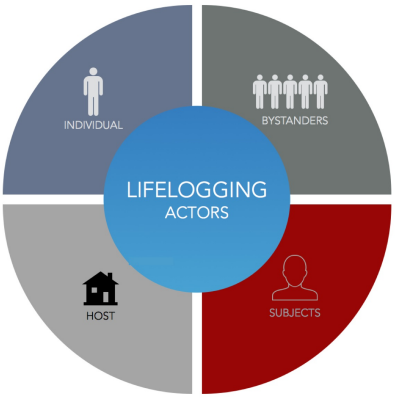

Figure 3. The four different actors involved in lifelogging

every image contained within that event (on average, over 100 images per event). It would have been unfeasible for the lifelogger to get the permission of every individual that they interacted with, so the lifelog is an absolute representation of the activities and interactions of the lifelogger. This browser does not attempt to implement any privacy policies, relying instead on the lifelogger to maintain sufficient control of the data and a sufficient awareness of the privacy implications of sharing or making public any of the data. Considering the current range of commercial lifelogging solutions (e.g. the Narrative Clip or the OMG Autographer), these all implement a similar approach to privacy, relying on the lifelogger to implement their own privacy policies.

2.3. Lifelogging Actors

Although we consider lifelogging to be a solipsistic activity in this article, the lifelogger is only one actor in the lifelogging universe. In reality, the lifelogger interacts with other individuals on a continual basis, so if we are to consider the privacy implications of these interactions, then it is important to identify the actors (interested parties) that we can identify in lifelogging[7] as illustrated in Figure 3.

- The Lifelogger. The individual who wears the devices and the individual who has a record of all life activities stored in a digital repository[8]. In most prior discussion, the privacy of the lifelogger was not considered, yet the lifelogger entrusts details of all life activities to the digital memory and the host service provider. A lifelog will inevitably capture aspects of the lifelogger's activities (for example, scenes A & B in Figure 4, credit card details, information accesses, interpersonal encounters) that could be harmful if made available publicly. Given the private nature of lifelog data, it is very likely that the lifelog of some individuals will be a target for unauthorised access attempts. It is our consideration that the lifelogger

[7] We identify four actors, but it may be reasonable to include society as a fifth actor because of the potential benefits that society can gain in terms of security and healthcare and the potential costs in terms of societal loss of privacy. However, we have chosen to focus the discussion on these four directly involved actors for this article.

[8] In prior work, the lifelogger does not need to be a human, it could just as easily be a robotic device [29], an environment or a real-world object.

Figure 4. Examples of potentially private data (A & B), Bystanders and Subjects (C & D) in a Google Glass captured lifelog from 2014.

should have the most privacy concerns, given the nature of lifelog data and given the lack of secure digital memory solutions as of yet.

- The Bystander. The people who are captured (inadvertently or on purpose) in the lifelog data stream of an individual. Depending on the lifestyle of the lifelogger, these could be predominantly strangers or they could be work colleagues, family members, friends, etc. The key factor about the bystander is that they are not the direct subject of the image. Consider the bystanders in Figure 4, C & D.
- The Subject. We consider the subject to be a unique type of bystander who is interacting with the lifelogger. Hence the subject will form a sizable percentage of the lifelog data (e.g. photos or sound) at any particular point in time. Examples of the subject can also be seen in 4, C & D.
- The Host. The lifelog needs to be stored and organised by some entity, referred to as the host. Perhaps the host is also the lifelogger, or perhaps it is some online web service that the lifelogger entrusts with the lifelog data. While the privacy of the host is not typically at risk, the host bears the responsibility for maintaining the privacy of the other three actors by keeping the data secure and by the implementation of appropriate privacy provisions.

Each of these actors are likely to see privacy differently and different forms of lifelogging are likely to have very different impacts. For example Figure 5 shows the different actors who would have privacy concerns across a scale of lifelogging scenarios, from basic quantified self-analysis to the extreme visual/aural lifelogging that we consider in this article. With regard to quantified self analytics, privacy is not such a big issue. A wearable pedometer or a heart-rate monitor with local storage pose few issues, even if the device is stolen. Progressing to a cloud-hosted technology, such as the Moves app, or other cloud-hosted fitness apps, the potential for privacy issues is dependent on the

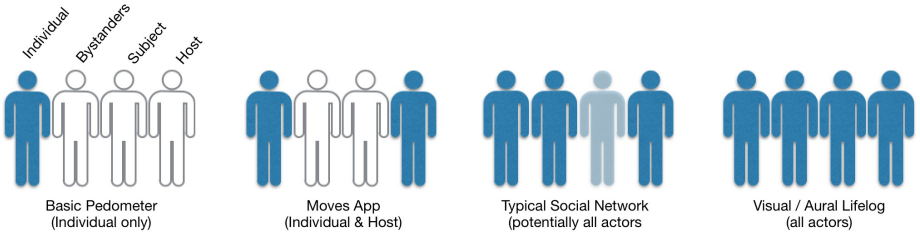

Figure 5. Privacy implications for different actors across various personal data gathering tools (individual, bystander, subject and host from l to r).

type of data stored; Moves, for example, stores ambient locations for the individual. Our current generation of social networking applications, such as Facebook, pose more privacy concerns, but once again, this is dependent on the type of information posted by the individual as well as on the efforts at securing the host server from malicious attack. At least in the case of social networks the individual usually has the ability to curate their own content and their being tagged in others' content. Finally, the rich visual/aural lifelog; this is the case in which the most privacy implications are noted. The privacy of the lifelogger, the bystanders and the subjects (person being spoken to/interacted with) are at risk due to the fact that a typical lifelog is uncurated.

In the case of an individual engaged in the extreme end of lifelogging, the lifelogger may be interested in maintaining detailed media-rich digital diaries while protecting their own privacy; the subjects may be concerned with their image being used inappropriately or published without permission; while the host would probably want to monetise the data in some way and may see privacy policies as an impediment to doing so. Hence we can see that the actors will all have different expectations and requirements and that the different forms of lifelogging involve different interactions between these actors.

In fact lifelogs have arrived in the world at a time when privacy has one again become a hot topic. There is genuine distrust of the corporations that hold most of our private data - organisations which did not even exist twenty years ago. If lifelogging is to become a mainstream activity, then the positive benefits should outweigh the privacy concerns, so we should explore some of the application scenarios for lifelogging.

2.4. Lifelogging Application Scenarios

There have been many use cases proposed for lifelogging technologies. In their book "Your Life Uploaded", Bell & Gemmell [11] suggest that lifelogging has the potential to revolutionise our healthcare by providing new sources of information about our bodies; improve our learning by providing new semantic tools to help replace memorisation with understanding; enhance our productivity by providing us with new reflective knowledge on our performance at work; and finally, improve our social lives by providing new life experience summarisation and review technologies.

At present, there is already a deep penetration of a basic-form of lifelogging; that of the quantified self movement [2]. There are already a large number of quantified-self applications which show successful inclusion of a basic form of lifelogging technologies and concepts. Many of these are based around some form of personalised healthcare or wellness monitoring, such as the range of products on the market which log caloric

energy expenditure and types of human physical activity being performed such as the FitBit One™which is worn as a clip-on device on the belt or trouser pocket; or the Nike FuelBand™, worn as a bracelet[9]. There are few privacy concerns with this data, unless the data is accessed in a malicious manner, and even then the privacy implications may not significant for the individual [10].

However, moving closer to the vision of capturing the totality of life activities, the Moves and SAGA smartphone apps capture all life activities (locations, activities) of the individual in a non-visual manner and present them in a basic version of a lifelog. For such apps, there is very little concern for privacy, save for the privacy of the lifelogger themselves, who may inadvertently lose or publicise too much personal data.

Using visual lifelogging, we can see applications for triggering recall of recent memories; an application of lifelogging where the detailed lifelog acts as a memory prosthesis, thereby providing support for people with Alzheimer's or other forms of dementia [31,32,33]. Extending this concept from the assistive/caregiving domain to every-day life, there is potential for lifelogging to provide memory support to fallible human memory [34]. For example, the lifelogger could be reminded of the last time they had similar interactions with the same people or objects. Any object in the real-world could be used as a real-time query to locate that (or similar) object (s) in the lifelogger's past.

There is also potential for lifelogging technologies to be used by organisations as a means of recording/logging the activities of employees for various reasons, such as keeping detailed records for legal/historical reasons, replacing manual record taking, logging information access activities as in [35], or potentially as a new technology to support aspects of organisational memory [36].

Lifestyle analysis to provide feedback - perhaps on dietary habits - is another possible use case. There has also been consideration of lifelogging as a self-protective technology. The sousveillance concept proposed by Steve Mann [3] is concerned with empowering the individual by the use of wearable computing and lifelogging technology. Mann's proposal is that a society that supports pervasive surveillance by authorities should also support sousveillance by the individual of their activities and surroundings.

So we can see that lifelogging may have a range of beneficial use cases, and we have only presented a small selection here. Many of the eventual use cases may not even be understood yet, in much the same way as the diverse uses of the WWW could not have been foreseen twenty years ago. Initial use cases (perhaps implemented on a technology such as Google Glass as a real-time digital memory search tool) may attract the first generation of early lifeloggers, but it is the new and novel uses of lifelogs that will follow which would probably, in turn attract additional lifeloggers, and a cycle of adoption will begin.

2.5. Why is privacy important for lifelogs?

The concept of privacy has been briefly discussed in the introduction. For lifelogging, there are some unique reasons why it is particularly important to understand the implications for privacy. Based on our eight years of experience with real-world lifelogging, we propose the following reasons why lifelogging would raise privacy concerns.

[9] http://www.fitbit.com, http://www.nike.com/fuelband

[10] This topic has been discussed in the press and various suggestions have emerged about the potential of fitness trackers to expose infidelity by analysing heart-rate correlations with pedometers and time-of-day [30].

1. Lifelogging as the norm. The continual progression of technology, cyclical innovation, and commercial interests in understanding real-world user context, suggests that lifelogging could become a more popular activity in the coming years, progressing from the early-adopter phase to wider population penetration.
2. Lifelogs are media rich. Images, audio, video, locations, activities, communications, information accesses; huge archives of data that can contain very personal information about three of the actors (individual, subject and bystanders). As technology progresses, it is likely that the richness of lifelog media will continue to increase, moving from photos and audio snippets into wearable video at HD or 4K resolution.
3. Lifelogs are uncurated. Anyone keeping a diary curates the data naturally. They probably choose not to record embarrassing facts or content that could potentially cause problems if accessed by others. With a lifelog, the content is automatically logged and stored; every meeting, every meal, every affair, every road traffic violation, potentially even every spoken word. It is reasonable to assume that the lifelogger will forget, be unable to find, or even be unaware of much of the data contained therein. As the lifelog becomes larger, stretching across years and decades, then this becomes a more extreme problem. Consider the challenge if a lifelogger were asked to remove any image or traceable audio of an individual they had spoken to a handful of times over a number of decades.
4. At present, lifelogging occurs indiscriminately. It would require an impossible level of curation to ensure that a lifelog did not contain the image, sound or any representation of another individual. Potentially, every single face stored in a lifelog could be detected and blurred beyond recognition, or every spoken word could be identified and masked. However, such a blunt instrument would result in some of the key benefits of lifelogging potentially being lost.
5. No permission to capture. No matter how well motivated the lifelogger is, it is essentially impossible to get the permission of everyone who appears in a lifelog[11].
6. All-of-lifelogs. Unlike security and CCTV camera installations that are generally tolerated because of the assumed benefits, as well as the expectation that the recorded data will be deleted after a number of days or weeks, *life*logs should be stored for a long time. Potentially, such lifelogs could grow to map the entire life of an individual, including every interaction that they have had.
7. Lifelogs are huge archives containing inter-dependent life activities. An isolated statement, image or sensor sequence could easily be made to give the wrong impression. Hence, a third party accessing a lifelog could construct stories and untruths based on lifelog data, or a short snippet could be presented out of context to cause harm to the lifelogger.
8. Lifelogs as targets. Given human nature, it is inevitable that lifelogs could be/will be targets for certain people who wish to to cause harm to the lifelogger. For various reasons; people with grudges, ex-partners, journalists, law enforcement, governments, employers, could attempt to 'hack into' the lifelog of an individual to seek information or to cause harm.

[11] Sitting here in a cafe writing this article, there are seven people directly in range of my lifelogging camera, with others passing continually by the window outside. It would not be feasible to expect the lifelogger to obtain the permission of all these individuals.

9. Security of lifelogs. The location of the lifelog needs to be considered and secured. It is unlikely that all but the most dedicated and knowledgeable lifelogger will be able to fully secure their own lifelog if they host it in their own home/on their own machines. If the lifelog is hosted on a server as part of some lifelog service provision, then the service must be trusted and must ensure the security and integrity of the data contained in the lifelog, as well as implementing privacy provisions over the data.

As can be seen from the above list, there are generally two categories of privacy concerns in lifelogging, those of the lifelogger and those of the people with whom the lifelogger interacts. The lifelogger would be primarily concerned with their own privacy; the lifelog might contain very personal information ranging from credit card details and health details to all interactions with other people and potentially intimate or embarrassing moments. Hence the privacy interests of the lifelogger focus around data protection and security. This is governed by the relationship between the lifelogger and the host. The lifelogging solution must be secure from end-to-end, otherwise the risk to a lifelogger will be too great and the benefits may not outweigh the significant personal cost of gathering the lifelog. As we have seen above, human nature suggests that a lifelog which is not securely stored may become a target for certain individuals who may wish to examine or explore the lifelog of another for various reasons. Since this is primarily a data-security issue, this will not be the focus of this article, rather we will focus on the privacy concerns of the bystanders and subjects.

When it comes to the people that the lifelogger interacts with, the primary concern would probably be how they are represented in the lifelog of another, over which they would have no control. They could be identified as being in a certain place at a certain time, engaged in activities, potentially with others, or being recorded saying certain things. Never before have individuals and societies had to address the potential of such pervasive recording. Widespread adoption of lifelogging - sousveillance, personal life-recording - whatever it may be called, would bring profound changes to the expectation of privacy, where the reliance is on hosts and lifeloggers to maintain lifelogs in a responsible manner. Hence it becomes important to put in place a framework that both respects the wishes of individuals to remain private (i.e. not to be identifiable in the lifelogs of others), yet allows for full lifelogging to take place. What we will present is a first version framework; an early framework that we are implementing as a testbed for a first generation of privacy by design lifelog. This framework focuses on private lifelogging and had been developed accordingly. Future versions of this framework will address the issue of lifelog sharing and publishing.

3. Privacy and Lifelogging

Human beings have been concerned with privacy from their earliest days. Whether by wearing clothes to protect personal modesty or the erection of fences and walls, humans have been protecting privacy for millennia. In recent years, the increased ability to gather and send information has generally had negative implications for retaining personal privacy. Many of the web services that we take for granted today are in existence due to our willingness to turn a blind-eye to their use of our interaction histories. Web search, social networking, email services - they are successful because we are willing to relin-

quish some aspect of our privacy in order to receive the service at low or zero cost[12]. In addition, we are also aware that our emails, instant messages and social networking updates are all likely to be stored indefinitely.

However, if the extreme forms of lifelogging progress to become a normal activity, there will be so much information stored in lifelog databases that an individual would have no practical means of controlling, or even knowing about, all of the visual and aural representations about themselves that others may hold or have access to. We are already experiencing this in terms of omnipresent surveillance cameras, but because of the perceived benefits, we are willing to ignore the potential of these cameras to track our movements though cities and environments. It may be that society will prove to be more willing to trust an unnamed organisation with a security camera than an individual engaged in lifelogging.

Before we begin to discuss what a privacy-aware lifelogging solution could look like, we need to understand what we mean by privacy and explore the potential impact of lifelogging on the privacy of individuals. As we have mentioned, the lifelogger is concerned with their own privacy because of the personal nature of the lifelog, but this is primarily a data security issue[13]. The bystanders and subjects may be concerned about being recorded/lifelogged by someone without their permission or without any notice being given. The host will need to assume responsibility for implementing whatever privacy policies are deemed necessary for hosting the lifelogs in a secure manner.

3.1. What do we mean by Privacy in terms of Lifelogging?

When considering how to develop lifelogging solutions, one of the first concepts that we need to understand is privacy and what a reasonable expectation of privacy is. Although it may come as a surprise to the reader it is not possible to find an agreed definition of privacy or an expectation of privacy. Privacy is an inherently fuzzy concept [37]. Historically it has had many meanings and definitions, and expectations of privacy differ across jurisdictions and eras. In the case of the opening quotation of this chapter, it was defined loosely by Warren and Brandeis in 1890, following Judge Thomas Cooley, as the *'right to be let alone'*. This is generally considered to be the first definition of privacy and came about as a result of the increasing availability of newspapers and photographs. However this is a vague definition and not one that can easily be applied to any modern day technology, let alone lifelogging[14].

Seeking a more recent definition, we look to 1960, where privacy has been defined in terms of a set of four torts[15][38], where a tort is a civil wrongdoing which causes

[12]However, one can still choose to use more privacy-preserving tools, such as the Duck-duck-go search engine, instead of Google.

[13]In much of the prior research, the concept of privacy is often confused with the concept of security of data. In this work, we clearly separate them out. Privacy of the lifelogger is concerned with keeping the data secure and under the control of the lifelogger. Privacy in terms of the individuals the lifelogger encounters is the focus of this article.

[14]After all, an individual can wear a 4K video camera in crowded public places, but without interacting with anyone and be said to be leaving them alone.

[15]A tort is generally considered to be a civil wrong which unfairly causes someone else to suffer loss or harm resulting in legal liability for the person who commits the tortuous act, called a tortfeasor. The four torts were; intruding (physically or otherwise) upon the solitude of another in a highly offencive manner, publicizing highly offensive private information about someone which is not of legitimate concern to the public, publicizing

unfair loss or harm to another. The working assumption is that privacy breaches are highly offensive or result in unfair loss being suffered by someone. However, neither what constitutes an offensive manner, nor how to quantify unfair loss are clear. Another potential definition comes from the Universal Declaration of Human Rights, in which a right to privacy is explicitly stated under Article 12, which speaks about privacy in terms of interference and attacks: *'No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks'*, but once again, it is unclear how a solipsistic lifelog can cause such problems, unless the lifelogger purposely and intentionally uses some subset of the lifelog to cause harm to another.

It turns out to be very difficult to define either privacy or the right to privacy, since privacy is not a pure legal term. We know from [39] that privacy has psychological, social and political aspects. The conceptual difficulties in defining the right to privacy are due to the fact that various interests protected by the right are also protected by other laws, and due to the fact that privacy is affected by political, social and economic changes and by technological developments. A definition of the right to privacy in a digital world, from [39] 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"*. In this definition, consideration is given to the individual having the right to choose what aspect of their domain can be accessed by others and how it can be used.

While ambiguities still inevitably arise and we cannot hope for one definition to cover all jurisdictions, we will adopt the spirit of this definition for our work. Privacy in terms of lifelogging and digital capture refers to both the privacy of the lifelogger and the privacy of the subjects and bystanders who may not wish to be identifiable. Taking this viewpoint, we consider that the right to privacy in terms of lifelogging refers to *the right to choose the composition and the usage of your lifelog and the right to choose what happens to your representation in the lifelogs of others*.

In an ideal situation, the digital memory host will protect the privacy of the lifelogger by implementing strong security measures[16] and protect the privacy of the subjects and bystanders by allowing them to define privacy policies that dictate whether they are in or out of any individuals lifelog at any point in time. So this will become our aim in defining a privacy-aware lifelogging framework; the assumption that every individual should be unidentifiable (image, audio or otherwise) unless they have given permission for a lifelogger to view their representation in that lifelogger's data.

However, how this simple assumption translates into digital memory software is unclear. Does it mean that a lifelogging device should somehow avoid capturing individuals into a lifelog as it is being gathered? We suggest not because such a blunt instrument

a highly offensive and false impression of another and using another's name or likeness for some advantage without the other's consent.

[16]Exactly what these strong security measures are is beyond the scope of this article, yet the concept of peer-to-peer military grade encryption could allow the company to host a lifelog, receiving lifelog data and metadata annotations, processing it locally and not actually have access themselves to the raw lifelog data. Essentially this renders all content undiscoverable by anyone except the lifelogger.

Figure 6. The five stages of data privacy consideration in lifelogging. The first three stages have no human involvement. The final two stages involve the data being accessible by humans.

(capture-time data removal) is a one way process and there would be no way to give retrospective permission for your representation to appear in someone's lifelog. In reality, there are five stages of data processing that can have an impact on privacy in lifelogging. These stages are shown in Figure 6.

The five stages range from physical data capture, through storage, processing in a lifelog and eventual access, with the automatic stages of the process on the left and stages which involve humans on the right:

- Capture. The physical capture of the raw pixels that may include the representation of an individual against their wishes. In some jurisdictions, simply sampling the pixels could be considered to be a breach of privacy, depending on the use cases of the data. For a long-term lifelogging scenario, it is unlikely that the lifelogger would even be aware of all the content that has been captured.
- Storage. The storage of these pixels in a computer storage device. Since storage would typically imply later usage[17], this could potentially cause problems in some jurisdictions. Given our definition of lifelogging, storage of the data is an essential component. Using computer vision or audio processing techniques, there is the potential to capture an image or sound sample, to process it immediately in order to extract the semantic meaning and then never actually store the sample. This might be necessary in some use cases, but the downside of this is that the data could never be re-analysed in a processing phase, and is thus lost to the possibility of applying new data processing methods (e.g. AI techniques) to enhance the utility of the lifelog.
- Processing. The automated lifelogging analytics carried out on these images, such as event segmentation, face/object detection - and potentially recognition. This process can either be automated or involve humans. Typically this is carried out on data that has been captured and stored. Over time, as new processing technologies become available, it will be important to re-process older lifelog data to extract additional valuable semantics.
- Access. A form of processing that directly involves the lifelogger accessing the lifelog using the digital memory software for personal (not publication) use; this is the point made earlier about lifelogging thus far being a solipsistic activity. Given that the previous steps are likely to be automated processing, this is the first phase of the process in which a human is likely to be involved. In this case,

[17] An interesting aspect of visual lifelogging and also a criticism that has been leveled at it is that a lifelog is primarily a WORN technology - Write Once Read Never. For the majority of a lifelog data, this is a reasonable assumption, however the view proposed by Bell and Gemmell in Total Recall[11] is that one never knows when some piece of data could become very valuable.

user access could take the form of quantified self-style aggregated data, or actual searching/browsing/viewing lifelog data as a memory support/reminiscence tool, as in Figure 2.

- Publication. Another form of processing involving a human in which the lifelogger (or potentially someone else) decides to make publicly available, via some means, some part of the lifelog, which may or may not contain recognisable representations of individuals. It is our opinion that this phase poses most privacy concerns as there is potential to cause harm to the bystanders or the subjects in the lifelog.

While we suggest that it may be impossible to define a one-size-fits-all-jurisdictions lifelogging solution, we propose that (referring again to Figure 6) the first three phases (Capture, Storage and Processing) pose a significantly reduced privacy threat to the bystanders and subjects, assuming the provision of an adequate level of data security. The most significant privacy concerns actually arise when human involvement takes place (Access and Publication). We propose that the digital memory should be capable of storing the compete representation of the life activities of the lifelogger indefinitely, but that the access and publication be restricted to protect the privacy of third parties. This is essentially a *caveat scriptor* policy, referring to the Latin for writer/publisher beware. If there is an identifiable tort as a result of publishing the representation of an individual, then that individual should have the right to remedy.

3.2. Prior Consideration of Privacy in Lifelogging

In previous discussions of lifelogging and privacy, the concept of privacy has been discussed, but rarely carefully considered. In developing working prototype lifelog solutions, the focus has been on developing the data organisation and access tools, or on the benefits in a particular use case. Concerning privacy, Allen [28] has discussed lifelogging and privacy from a legal viewpoint and suggests that no one should record or photograph others for a lifelog without the consent of the person or their legal guardian, and has gone so far as to propose that a counter-technology to block lifelog surveillance should be developed. However, no guidance on how these suggestions could be implemented is given.

In other work, the Privacy Butler [40] is an automated service that can monitor a person's online presence and attempt to make corrections based on policies specified by the owner of the online presence. It is a tool with which the subjects and bystanders can explore their representation in the public data of others. Zhou and Gurrin, in a survey of lifelogging technologies and the attitudes of people towards lifelogging [41] identify three primary concerns of individuals: privacy, appearance and comfort. In this work, privacy was considered important, both in terms of the lifelogger and the individuals that the lifelogger encountered.

When developing a set of ethical guidelines for employing lifelogging for research studies, Kelly et al. refer to privacy in terms of gathered image data that could be intrusive, potentially unflattering and unwanted [42]. They take a health-care researcher's viewpoint and define a listing of guidelines that help to preserve the privacy of lifeloggers and bystanders/subjects in controlled user studies. One of the key guidelines from this work is concerned with the publishing of data, and we are told that 'the privacy and anonymity of third parties must be protected; no image that identifies them should be

published without their consent'. What is not described is where the balance lies between the rights of the lifelogger to gather their own data and the rights of the bystanders and subjects to be left alone.

The most considered discussion of privacy and lifelogging comes from O'Hara et al. [9] who look at the technological possibilities and constraints for lifelogging tools, and set out some of the most important privacy, identity and empowerment-related issues. As lifelogging becomes more mainstream, the authors point out that the privacy concerns of society may be offset by the empowerment of the individual as new lifelogging applications come on-stream, or we may develop lifelogging systems that integrate 'privacy by design' into the development process.

3.3. Privacy by Design

Privacy by design is a proposed framework for ubiquitous computing [43] in which privacy and data protection are embedded as core considerations throughout the entire life cycle of a technology, from the early design stage to their deployment, use and eventual disposal. Privacy by design principles are based on seven foundational principles [44]: proactive not reactive; privacy as the default configuration; privacy embedded into the design; privacy as additional (not reduced) functionality; end-to-end data security; visibility/transparency; and respect for the privacy of the individual user. These seven principles provide an initial set of guidelines for developing privacy-aware lifelogging systems. While privacy by design is obviously a positive concept, it has received criticism for being vague and lacking detail about how to actually implement such a concept while meeting the functional requirements of the system under development; for details see [45]. With regard to lifelogging, there is an inevitable trade-off between privacy and functionality, and where lifelogging settles on this trade-off scale is yet to be seen. In any case, the concept of privacy by design is likely to become a core component of any large-scale lifelogging solution and it is something that software architects and developers should take into account when developing lifelogging organisation and retrieval tools.

Sipekermann [37] proposes a number of challenges for implementing privacy by design solutions[18], including: the fuzzyness of the privacy concept, the lack of an agreed-upon development methodology, and a lack of knowledge about the impacts of privacy practices. Hence, we can see that privacy by design is not a perfect set of principles, but gives us guidance on how to implement a lifelogging framework.

3.4. A Privacy Aware Lifelogging Framework

Considering the discussion to date, our experiences of lifelogging over eight years, our understanding of the technical feasibility of our suggestions, our understanding of the concept of privacy and the guidelines from privacy by design, we propose an initial privacy by design lifelogging framework which could potentially be integrated into a fully functional digital memory in the spirit of Bell and Gemmel's Total Recall vision [11]. The key aspects of this framework are as follows.

[18]There are a number of challenges proposed, but we mention the relevant ones only. For a full review, please see [37]

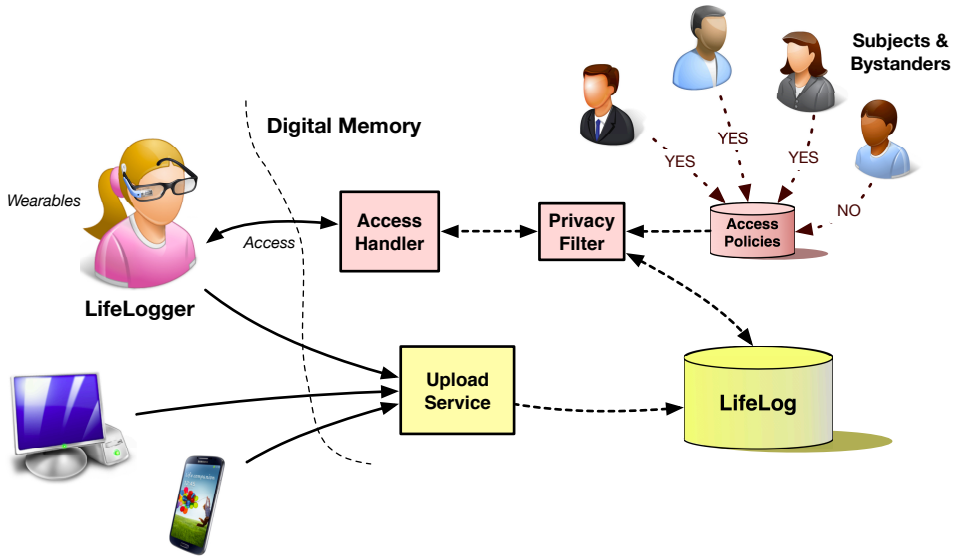

Figure 7. A summary lifelog architecture implementing principles of privacy by design.

- High fidelity automated capture of the totality of life experience, with no privacy-based limitations on data capture[19].
- Secure transmission of lifelog data from the wearable devices to the hosting service (digital memory) and onward to the lifelogger when required.
- Secure hosting of the lifelog data, with access rights for the lifelogger, governed by privacy policies.
- Sufficient level of data analytics to extract adequate semantics from the lifelog data[20] to support the use cases of the lifelog.
- Implementing a privacy by design approach to lifelog access whereby the privacy of all actors involved is considered and secured (in our prototype by means of face-detection and blurring).
- Anyone has the right to choose to be in/out of another's lifelog, in the same manner as anyone has the right to partake in lifelogging themselves or not.

Following the criteria of privacy by design, the framework attempts to balance the need for privacy with the maintenance of a data-rich lifelog to support a wide range of potential use cases. An architecture overview is shown in Figure 7. The lifelogger gathers data in an indiscriminate manner and this data is securely transmitted to the digital memory for processing. The digital memory stores all the lifelog data in its original

[19]Exactly how much data is captured and in what format is not of concern here. We assume that the data (video/audio and image currently) is enough to capture a detailed life trace for the lifelogger and sufficient to arouse privacy concerns.

[20]For more consideration on the semantics of lifelog data see[4].

form[21] and will regulate access by means of privacy policies[22] that are maintained for every bystander that the system has identified and knows about. Taking this approach, the default privacy policy does not allow the digital memory to display any recognisable representation of an individual (in this case, a face), unless that individual has given permission. These privacy policies regulate access via *dynamic views* over the lifelog and reflect the current set of privacy policies. These dynamic views are implemented by the digital memory (Privacy Filter in Figure 7) and in this way, any individual is free to choose which lifelogs their recognisable image can appear in, because the system assumes privacy as the default, i.e. every face is blurred unless the face detection algorithms identify it as a known face. We refer to this as 'negative face blurring', because we blur every face except the ones that we can identify as 'ok' for our lifelog.

In their most basic form, these policies include a sequence of face models for every person who has given permission to appear in the lifelog. Typically, one would imagine that there would be a small number (100 or so) of such face models, probably comprising the family and close friends of the lifelogger, or it could perhaps be based on an existing social network. The assumption underlying this framework is that the privacy preserving implementation needs to be flexible to changing privacy requirements, both in terms of an individual removing the right to their image, or reinstating the right at a later point. Consequently, referring back to Figure 6, the lifelog data (camera and other sensors) is captured in a normal manner, it is stored permanently in the lifelog and the digital memory processes the data to identify known faces; any privacy filtering occurs during the Access and Publication phases.

Since we separate the *view* of the lifelog data from the actual stored data, then these policies can be updated in real-time so that an individual can retrospectively add or remove access rights to their identifiable image. We call this approach *dynamic policy-driven negative face blurring*.

We have developed a prototype [46] of the first version of this framework (shown in Figure 8) and are working to improve and expand it. In this screenshot, a subject is seen who has not given permission to the lifelogger, so the face is detected and automatically blurred. The proposed framework utilises face detection and face recognition as underlying technologies to implement a privacy preserving policy and implements it in real-time for a Google Glass gathered lifelog. Performance and accuracy of the face detection are in line with state-of-the-art face detection approaches. The basic lifelogging capture and storage components of this prototype are based on pre-existing solutions developed previously [13]. With regard to processing, the key challenge is twofold; firstly to find all potentially recognisable faces in every lifelog image (or video) and secondly, to identify the subset of faces that are allowed to be identifiable based on the privacy policies of these individuals. Face detection and recognition algorithms are continually improving, and initial results of negative face blurring are discussed in [46]; the benefit of this approach is that if a face cannot be identified, then it is blurred by default. One can expect

[21] If lifelogs become a self-protection and defence mechanism, it is likely that the lifelogger will need to be able to prove that the original lifelog data was never tampered with. Kelly et al. also point out that 'the law may not permit privacy', explaining that if the images depict any illegal activities, according to national regulation, the researcher may be under a legal and professional obligation to breach confidentiality and pass on image data to appropriate authorities.

[22] Exactly how the privacy policies are to be implemented is a subject that we are looking into at present. There may be a privacy policy look-up service, or the individual lifelogs may maintain a simple privacy policy for each willing bystander and subject.

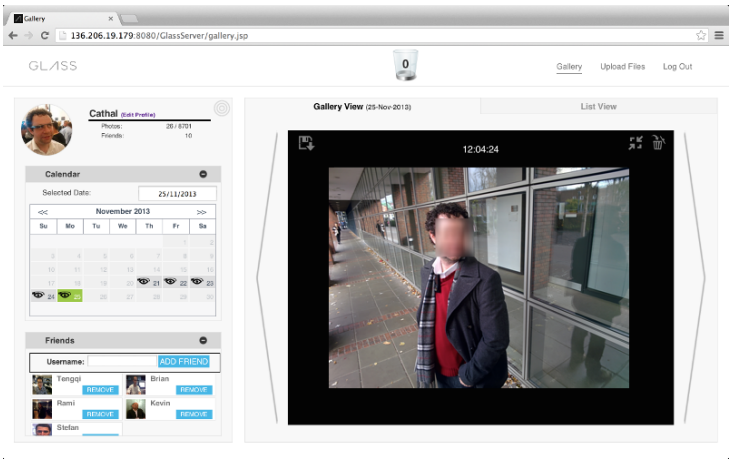

Figure 8. The first implementation of a privacy by design lifelogging system.

the accuracy of face detection and recognition algorithms to continue to improve and to be refined to exploit artifacts of lifelog data, such as individual co-location or enhanced sensing.

Another challenge is how to implement a fluid privacy policy infrastructure that is updatable and accessible in real-time. We consider this to be a software engineering issue and one that is inherently solvable using readily-available technologies. Hence, although the proposed framework is a prototype, the technologies required to make it available as an online service are already available. Whether such a framework emerges naturally as a lifelogging software service remains to be seen. None of the current embryonic lifelogging solutions implement such a framework.

4. Comment and Conclusion

In this chapter, we have motivated and presented a first privacy-aware lifelogging solution that allows for indiscriminate lifelogging, but attempts to protect the privacy of individuals involved. It is our belief that the individual has the right to choose what happens to their own lifelog and the right to choose what happens to their representation in the lifelogs of others.

Lifelogging is a new technology and how it progresses has yet to be seen. The realisable benefits of lifelogging are not yet understood, but we believe that it has the potential to be a life-changing technology. However, we may find that society views indiscriminate, rich lifelogging as a step too far, or there could be legal impediments to lifelogging. It is possible that privacy laws or societal demands will result in a lifelog privacy framework becoming a necessity and that all lifelogging service providers will have to sign-up to global privacy-preserving data access policies. The discussion around such issues is only just beginning and will end up shaping the types of lifelogs that are maintained in the future.

However, like O'Hara et al. in [9], it is our consideration that the benefits of lifelogging will probably outweigh the costs. There is a scale of lifelogging, from basic quantified-self analysis to full indiscriminate media-rich lifelogging, and the point on

that scale where the lifelogging occurs has major implications for the privacy of those involved. The point on that scale society settles on a tolerable level of lifelogging has yet to be seen.

In reality, the use cases will have a major impact on the acceptance or otherwise of lifelogging. Should a memory-impaired individual who relies on a lifelog as a surrogate memory be denied the ability to record, store and refer to a lifelog on the basis that it could potentially breach the expectation of privacy of an unknown individual, even though maintaining a lifelog provides a significant aid to human memory, as in the case of Mrs B, one of the first SenseCam wearers ever studied[47]? What about the (reasonable) scenario that a digital memory (lifelog and software) can provide immediate access to a perfect memory for any individual, thereby enhancing life, productivity and learning, as in the vision presented earlier by Bell and Gemmell [11]? Or even more challenging, how about a scenario in which a caregiver can check on the status and whereabouts of a dementia sufferer who can consequentially enjoy freedom of mobility as a result of wearing a technology such as Google Glass with a lifelog sharing app? These are only three use cases of a technology that promises so much.

Naturally, privacy is still a major impediment to lifelogging uptake, as well as to actual research into the benefits of lifelogging. We cannot solve all privacy issues in all jurisdictions with one lifelogging solution, however, what we can do is look to best practice and take a flexible and sensible approach to securing lifelog data to protect the lifelogger and other actors by ensuring that the digital memory implement privacy policies. In this way, we can try to ensure the anonymity of unwilling (we mean people who have not actually agreed) bystanders and subjects, but still maintain a flexible digital memory which would be able to retrospectively apply new privacy policies without data loss or loss of integrity. We have taken the first steps in this process in the research presented in this article; however we are fully aware that there are enormous challenges to overcome. We have not mentioned other types of information that may identify people in lifelogs, such as voice, co-location, and so on. This is referred to by [48] as 'personally identifiable information' and this would need to be addressed in a working, real-world solution. In addition, we have not considered the (significant) search challenges, the data organisation challenges and the data presentation challenges [4] that will need to be addressed before lifelogging can begin to reach its potential. Neither have we considered the reasonable scenario in which an individual could enact a privacy policy that allows their representation to appear in the lifelog of any lifelogger.

Furthermore, for some use cases of lifelogging, rich media capture is not needed at all; for example, many of the quantified self activities. For other use cases, we are not interested in permanent storage because the data can be processed by algorithms immediately on capture, the semantics stored and the original data deleted. However, for a truly Total Recall [11] lifelogging vision, one would need to store the rich multimedia lifelog indefinitely.

We will finish by considering the precedence of how society has already accepted the digital recording of environments. Buildings and cars are allowed to 'wear' cameras in most jurisdictions and we rarely even consider this. If a lifelog can be seen as a form of self-protection (Mann's sousveillance argument), then why would society allow more freedom to protect the welfare of sweets on a department store shelf, than to protect the rights of a lifelogger to record for personal or medical use? We are at a turning point in human history; lifelogs are just one technology that has the potential to change what it

means to be human [11]. Other impending technologies, such as artificial intelligence, ubiquitous robotics, life extension, and nanotechnology all have the potential to impact hugely on our lives and in the same way as lifelogging, each one will face legal and ethical issues as they gain broader acceptance.

Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant numbers 07/SK/I/1186, 11/RFP1/CMS/3283 and SFI/12/RC/2289. The corresponding author wishes to express his appreciation to the University of Tsukuba for hosting his extended visit which took place in 2014.

References

[1] Samuel D. Warren and Louis D. Brandeis. The right to privacy. *Harward Law Review*, 4(5):193–220, December 1890.

[2] Jochen Meyer, Steven Simske, Katie A. Siek, Cathal G. Gurrin, and Hermie Hermens. Beyond quantified self: Data for wellbeing. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 95–98, New York, NY, USA, 2014. ACM.

[3] Steve Mann. Continuous lifelong capture of personal experience with EyeTap. In *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, CARPE'04, pages 1–21, New York, NY, USA, 2004. ACM.

[4] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. Lifelogging: Personal big data. *Foundations and Trends® in Information Retrieval*, 8(1):1–125, 2014.

[5] Gordon Bell and Jim Gemmell. A digital life. *Scientific American*, 296:58–65, February 2007.

[6] Steve Mann. Continuous lifelong capture of personal experience with EyeTap. In *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, CARPE'04, pages 1–21, New York, NY, USA, 2004. ACM.

[7] Niamh Caprani, Noel E. O'Connor, and Cathal Gurrin. Experiencing sensecam: A case study interview exploring seven years living with a wearable camera. In *Proceedings of the 4th International SenseCam & Pervasive Imaging Conference*, SenseCam '13, pages 52–59, New York, NY, USA, 2013. ACM.

[8] Kiyoharu Aizawa, Tetsuro Hori, Shinya Kawasaki, and Takayuki Ishikawa. Capture and efficient retrieval of life log. In *Proceedings of the Pervasive Workshop on Memory and Sharing of Experiences*, pages 15–20, Linz/Vienna, Austria, 2004.

[9] Kieron O'Hara, Mischa Tuffield, and Nigel Shadbolt. Lifelogging: Privacy and empowerment with memories for life. In *Identity in the Information Society*, volume 1. 2009.

[10] Martin Dodge and Rob Kitchin. "Outlines of a world coming into existence": Pervasive computing and the ethics of forgetting. *Environment and Planning B*, 34(3):431–445, 2007.

[11] Gordon Bell and Jim Gemmell. *Total Recall: How the E-Memory Revolution Will Change Everything*. Penguin Books, 2009.

[12] Dirk De Jager, Alex L Wood, Geoff V Merrett, Bashir M Al-Hashimi, Kieron O'Hara, Nigel R Shadbolt, and Wendy Hall. A low-power, distributed, pervasive healthcare system for supporting memory. In *Proceedings of the First ACM MobiHoc Workshop on Pervasive Wireless Healthcare*, page 5. ACM, may 2011.

[13] Zhengwei Qiu, Cathal Gurrin, AidenR. Doherty, and AlanF. Smeaton. A real-time life experience logging tool. In Klaus Schoeffmann, Bernard Merialdo, AlexanderG. Hauptmann, Chong-Wah Ngo, Yiannis Andreopoulos, and Christian Breiteneder, editors, *Advances in Multimedia Modeling*, volume 7131 of *Lecture Notes in Computer Science*, pages 636–638. Springer Berlin Heidelberg, 2012.

[14] Ray Kurzweil. The law of accelerating returns. In Christof Teuscher, editor, *Alan Turing: Life and Legacy of a Great Thinker*, pages 381–416. Springer Berlin Heidelberg, 2004.

[15] R.R. Schaller. Moore's law: past, present and future. *Spectrum, IEEE*, 34(6):52–59, Jun 1997.

[16] Walter Chip. Kryder's law. *Scientific American*, pages 7–25, 2005.

[17] J. Machajdik, Allan Hanbury, A. Garz, and R. Sablatnig. Affective computing for wearable diary and lifelogging systems: An overview. In Heinz Mayer, Martina Uray, and Harald Ganster, editors, *Machine Vision-Research for High Quality Processes and Products-35th Workshop of the Austrian Association for Pattern Recognition*. Austrian Computer Society, 2011.

[18] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. SenseCam: A retrospective memory aid. In *UbiComp: 8th International Conference on Ubiquitous Computing*, volume 4602 of *LNCS*, pages 177–193, Berlin, Heidelberg, 2006. Springer.

[19] Cathal Gurrin, Alan F. Smeaton, Daragh Byrne, Neil O'Hare, Gareth J.F. Jones, and Noel E. O'Connor. An examination of a large visual lifelog. In *AIRS 2008 - Asia Information Retrieval Symposium*, 2008.

[20] Stuart Dredge (The Observer). 10 things you need to know about – lifelogging, February 2014.

[21] M. Shah, B. Mears, C. Chakrabarti, and A. Spanias. Lifelogging: Archival and retrieval of continuously recorded audio using wearable devices. In *Emerging Signal Processing Applications (ESPA), 2012 IEEE International Conference on*, pages 99–102, 2012.

[22] Melanie Swan. The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data*, 1(2):85–89, June 2013.

[23] Susan Dumais, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin, and Daniel C. Robbins. Stuff I've seen: a system for personal information retrieval and re-use. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 72–79, ACM, 2003. New York, NY, USA.

[24] Mathieu d'Aquin, Salman Elahi, and Enrico Motta. Personal monitoring of web information exchange: Towards web lifelogging. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, 2010.

[25] David Elsweiler and Ian Ruthven. Towards task-based personal information management evaluations. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 23–30, Amsterdam, The Netherlands, 2007. ACM.

[26] Gamhewage C. de Silva, Toshihiko Yamasaki, and Kiyoharu Aizawa. An interactive multimedia diary for the home. *Computer*, 40(5):52–59, May 2007.

[27] A R Doherty, K Pauly-Takacs, N Caprani, C Gurrin, C J A Moulin, N E O'Connor, and A F Smeaton. Experiences of Aiding Autobiographical Memory Using the SenseCam. *Human-Computer Interaction*, 27(1-2):151–174, 2012.

[28] A.L. Allen. *Dredging-up the Past: Lifelogging, Memory and Surveillance*. Scholarship at Penn Law. University of Pennsylvania, Law School, 2007.

[29] Cathal Gurrin, Hyowon Lee, and Jer Hayes. iforgot: a model of forgetting in robotic memories. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, HRI '10, pages 93–94, Piscataway, NJ, USA, 2010. IEEE Press.

[30] Gregory Ferenstein. How health trackers could reduce sexual infidelity, July 2013.

[31] Steve Hodges, Emma Berry, and Ken Wood. SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory*, 7(19):685–696, 2011.

[32] Matthew L. Lee and Anind K. Dey. Using lifelogging to support recollection for people with episodic memory impairment and their caregivers. In *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, HealthNet '08, pages 14:1–14:3, New York, NY, USA, 2008. ACM.

[33] Ron M. Baecker, Elsa Marziali, Sarah Chatland, Kante Easley, Masashi Crete, and Martin Yeung. Multimedia biographies for individuals with alzheimer's disease and their families. In *2nd International Conference on Technology and Aging*, 2007.

[34] Aiden R. Doherty, Chris J.A. Moulin, and Alan F. Smeaton. Automatically assisting human memory: A SenseCam browser. *Memory*, 7(19):785–795, 2011.

[35] Sanna Kumpulainen, Kalervo Järvelin, Sami Serola, Aiden Roger Doherty, Alan F Smeaton, Daragh Byrne, and Gareth J F Jones. Data collection methods for task-based information access in molecular medicine. pages 1–10, 2009.

[36] Eric W Stein. Organization memory: Review of concepts and recommendations for management. *International Journal of Information Management*, 15(1):17–32, 1995.

[37] Sarah Spiekermann. The challenges of privacy by design. *Commun. ACM*, 55(7):38–40, July 2012.

[38] William L Prosser. Privacy. *California Law Review*, 48(1):383–389.

[39] Workshop S.I.L.T. *Privacy in the Digital Environment*. Haifa Center of Law & Technology.

[40] R. Wishart, D. Corapi, A Madhavapeddy, and M. Sloman. Privacy butler: A personal privacy rights manager for online presence. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on*, pages 672–677, March 2010.

[41] Lijuan Marissa Zhou and Cathal Gurrin. A survey on life logging data capturing. In *SenseCam 2012 - The proceedings of the 3rd Sensecam conference*, April 2012.

[42] P. Kelly, S. J. Marshall, H. Badland, J. Kerr, M. Oliver, and A. Doherty. An Ethical Framework for Automated, Wearable Cameras in Health Behavior Research. *American journal of preventive medicine*, 44(3):314–319, March 2013.

[43] Marc Langheinrich. Privacy by Design — Principles of Privacy-Aware Ubiquitous Systems. In GregoryD. Abowd, Barry Brumitt, and Steven Shafer, editors, *Ubicomp 2001: Ubiquitous Computing*, volume 2201 of *Lecture Notes in Computer Science*, pages 273–291. Springer Berlin Heidelberg, 2001.

[44] Ann Cavoukian. Privacy by design: The 7 foundational principles. implementation and mapping of fair information practices. *Information and Privacy Commissioner of Ontario, Canada*, 2010.

[45] Seda Gürses, Carmela Troncoso, and Claudia Diaz. Engineering privacy by design. *Computers, Privacy & Data Protection*, 2011.

[46] TengQi Ye, Brian Moynagh, Rami Albatal, and Cathal 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 knowledge management (CIKM '14). ACM, Shanghai, China (to appear)*, November 2014.

[47] Emma Berry, Narinder Kapur, Lyndsay Williams, Steve Hodges, Peter Watson, Gavin Smyth, James Srinivasan, Reg Smith, Barbara Wilson, and Ken Wood. The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis. *Neuropsychological Rehabilitation*, 17(4):582–601, 2007.

[48] Paul M Schwartz and Daniel J Solove. The pii problem: Privacy and a new concept of personally identifiable information,(2011). *New York University Law Review*, 86:1814.