

Lifelogging: Issues of Identity and Privacy with Memories for Life

Kieron O'Hara, Mischa M. Tuffield & Nigel Shadbolt

Intelligence, Agents, Multimedia Group

School of Electronics and Computer Science

University of Southampton

Highfield

Southampton SO17 1BJ

United Kingdom

{kmo,mmt04r,nrs}@ecs.soton.ac.uk

+44 23 8059 3523

Abstract: The growth of information acquisition, storage and retrieval capacity has led to the development of the practice of *lifelogging*, the indiscriminating collection of information concerning one's life and behaviour. There are potential problems in this practice, but equally it could be empowering for the individual, and provide a new locus for the construction of an online identity. In this paper we look at the technological possibilities and constraints for lifelogging tools, and set out some of the most important privacy, identity and empowerment-related issues. We argue that some of the privacy concerns are overblown, and the major issues will be concerned with surveillance. We also argue that much research and commentary on lifelogging has made the unrealistic assumption that the information gathered is for private use, whereas, in a more socially-networked online world, much of it will have public functions and will be voluntarily released into the public domain.

Key words: lifelogging, Memories for Life, Personal Information Management, privacy, identity, surveillance, sousveillance.

Introduction

The growth of computing capacity as predicted (and driven) by Moore's Law has meant that there are few barriers to the storage of information. In particular, a person can now store significant quantities of information about him- or herself. It has been estimated that, in the normal course of events during a life of normal duration, transactions involving a person will create something of the order of 100 gigabytes of information as a by-product. Gordon Bell and Jim Gemmell have estimated that a terabyte of storage would hold all the books, emails, recorded conversations, music tracks and photographs that one is likely to accumulate over sixty years, which a typical desktop PC will probably be able to hold by 2010, and personal digital assistants (PDAs) will manage by 2015 (Bell & Gemmell 2007). Alan Dix has noted that even 70 years of high-quality video recording would require something less than 30 terabytes of storage (equivalent to under 375 of Apple's largest iPods, which store 80 gigabytes and cost in the order of £150 each – cf. Dix 2002, O'Hara et al 2006).

These technological possibilities have led to the development of so-called *lifelogging* technologies and tools, to support the practice of exploiting digital storage systems to record information about a person, or group of people, automatically and persistently. Typical types of information to be logged include emails, documents, digital photographs and video, diaries/calendars, geodata using the Global Positioning System (GPS), music downloads, listening habits, blog entries and Web browser bookmarks and navigation history. The result for the user is a large store of information much of which will be trivial or ephemeral.

The purpose of storage of such information can vary, and may not be clear even to the lifelogger at the point of storage. However, in an information-intensive age where the surrender of digital identity is a commonplace, for purposes ranging from commerce, marketing, social networking, government, receipt of services, travel or security, lifelogging has the potential to reaffirm the individual's control of his or her own identity. The lifelog is a constructed identity that outweighs the others simply by weight of evidence, complexity and comprehensiveness. It is likely to include other identities, and amalgamate and supplement them.

This paper will examine the technologies behind lifelogging, and discuss the concerns about privately held 'Memories for Life'. It is clear that there are likely to be privacy concerns about lifelogging practice and technology, and indeed these concerns are already being raised (Allen 2008). The structure of the paper is as follows. The next section will set out some of the principles of lifelogging, and profile some of the more prominent efforts in this space. Then we will examine pragmatic aspects, such as the sorts of information that lifelogging is likely to draw upon, the uses to which it might be put and so on. Next, we shall examine some of the issues surrounding privacy, identity and empowerment for the lifelogger. Finally, we will conclude with a brief discussion of the arguments, noting in passing that personal knowledge management, in a time when increasingly many of our transactions (in the Western democracies at least) are migrating online, is fast becoming an imperative even for non-lifeloggers.

Lifelogging

Human memory is strangely unreliable. One remembers pointless stuff while forgetting someone's name, who won the cup last year, how to make coq au vin,

someone's telephone number, that one has a dental appointment at three, where one has put one's car keys. Much forgetting is particularly linked to memory function, so junking duff, unused or out of date information is good housekeeping for the head (cf. Schacter 2001), but unfortunately in the nature of the case it is impossible to forget all and only those things that one does not need.

Technology has been used to aid memory since, er, time immemorial. In Plato's *Phaedrus*, Socrates worries about the effects that the new literacy will have on Athenians' memories. But in the 21st century computing power has provided us with the means for augmenting our memories to a degree hardly dreamt of even ten years ago. We can store information in enormous quantities, so we no longer have to be selective. Removing the requirement for selectivity makes storage much less heavy in its use of other resources; e.g. one does not have to view it all to decide which to keep and which to delete. Gadgets such as cheap, small sensors make it increasingly simple to extract large quantities of data automatically from the environment. Increasing use of digital technologies means that many records of meetings or communications are in digital form already. And better search, retrieval and mining techniques mean that we are better able find the important signals in noisy data (O'Hara & Shadbolt 2008, 81-108).

We are getting to the stage when all information, interesting or otherwise, generated in a lifetime by a single person can be assembled in a giant autobiographical silo, and queried relatively efficiently, creating a need for *Personal Information Management* (PIM – Jones 2008). This can be seen as a step-change in the relation between the technology of information storage and human memory, and under the title 'Memories for Life' (M4L) has been adopted by the United Kingdom Computing Research

Committee (UKCRC) as a ‘grand challenge’ for computing research.¹ Memories for Life demands interdisciplinary research not only on technological infrastructure but also on social, legal, political, philosophical, psychological and medical aspects of the technological development.²

But why would we want M4L? Why should anyone want to keep emails whose point, if any, vanished the moment the computer was shut down? Surely there is a limit. There is surprisingly little research on memory prostheses, although Kalnikaitė and Whittaker (2007) have made some interesting discoveries of when exactly people are tempted to reach for the PDA rather than try to remember something unaided. Some people with severe memory impairments may rely on prostheses, and important work is appearing to show that patients with Alzheimer’s can use automatically-created photographs of apparently trivial events in their recent past to boost their short-term memories more effectively than other methods such as keeping meticulous diaries (Berry et al 2007).

A compelling reason often overlooked is the issue of association. Memory links are often associative – one is drawn to one topic by links to another, and recall is enhanced if background conditions can be recreated. So, for instance, suppose you were looking for a particular document on your computer without noticeable success. If you remembered that when you read it last you were listening to Mozart’s Clarinet Quintet on your MP3 player, and if you had stored information from the player and a properly integrated interface to all your personal data, you could restrict the search on the desktop to all documents that were open at the times at which the Clarinet Quintet

¹ <http://www.bcs.org/server.php?show=ConWebDoc.5203>.

² <http://www.memoriesforlife.org/>.

was playing on the player. There is nothing fundamentally interesting about the times at which particular pieces are played on the player, but there may be wider application. Information clarifying the context of an event – answering *what*, *where*, *when* and *who* – are key to supporting tasks of recall.

Fundamentally, there is a simple two-part positive reason for collecting all this informational clutter. First, we can do it because we can. And second, we can never know what will be useful in the future. Any piece of information might have value. Every piece of information is such that it is very unlikely, but just possible, that it is valuable. Before technology allowed comprehensive storage, our strategy was usually to try to estimate which information is likely to be more valuable and to keep that. Now there is no reason to stick to that philosophy.

As more of our transactions take place remotely, the need for our physical presence has declined, and increasingly many interactions are carried out by digital avatars of ourselves, in a general trend termed by sociologists ‘the disappearance of the body’ (O’Hara & Shadbolt 2008, 1-24). Hence our daily lives leave behind more evidence that we can collect and curate, and the indiscriminating collection of such evidence is what we call *lifelogging*. Lifelogging can be passive – one stores the by-products of the life one would have lived anyway – or active – one surrounds oneself with sensors and information capture tools to create as rich a picture of one’s life as possible.

Lifelogging has been the subject of investigation as long as it has been clear that digital technologies were going to be playing a large part in our lives. Steve Mann has spent a long time investigating wearable computing, in particular wireless cameras to record his daily existence, and has enjoyed media status as ‘the world’s first cyborg’

(Mann & Niedzviecki 2001).³ Jennifer Ringley also achieved a level of celebrity and indeed notoriety when she set up JenniCam, a webcam that recorded events in her living space, and made available on the Internet images of events in her daily life ranging from the mundane to the pornographic (cf. Jimroglou 1999). Initially she filtered out very private moments, while also sometimes ‘performing’ for the camera; but for most of the experiment, which ran from 1996-2003 the images were unfiltered. At its height, JenniCam attracted four million viewers daily.⁴

One important effort at lifelogging was LifeLog, sponsored by the American Defense and Advanced Research Projects Agency (DARPA), which became somewhat notorious, and which will be discussed later in this paper. Others include Total Recall⁵ (Cheng et al 2004) and SemanticLIFE⁶ (Ahmed et al 2004). The most famous and comprehensive lifelogging experiment is the attempt by Gordon Bell of Microsoft to create a digital archive, MyLifeBits.⁷

MyLifeBits ... has provided some of the tools needed to compile a lifelong digital archive. We have found that digital memories allow one to vividly relive an event with sounds and images, enhancing personal reflection in much the same way that the Internet has aided scientific investigations. Every word one has ever read, whether in an e-mail, an electronic document or on a Web site, can be found again with just a few keystrokes. Computers can analyze digital memories to help with time management, pointing out when you are not spending enough time on your highest priorities. Your locations can be logged

³ <http://wecam.org/>.

⁴ <http://www.arttech.ab.ca/pbrown/jenni/jenni.html>.

⁵ <http://bourbon.usc.edu/iml/recall/>.

⁶ <http://storm.ifs.tuwien.ac.at/>.

⁷ <http://research.microsoft.com/barc/mediapresence/MyLifeBits.aspx>.

at regular intervals, producing animated maps that trace your peregrinations. Perhaps most important, digital memories can enable people to tell their life stories to their descendants in a compelling, detailed fashion that until now has been reserved solely for the rich and famous. (Bell & Gemmell 2007, 40-42)

MyLifeBits began in 2001, but its roots were in Bell's attempt from 1998 onwards to go completely paperless, and scanned all documents (including logos on freebie conference mugs and t-shirts), videoed lectures, voice tapes etc, a high-cost strategy that required a full-time personal assistant. MyLifeBits was then set up to enable Bell to make sense of the enormous repository of information – in other words, to support querying, retrieval and search. MyLifeBits also provides new capture tools, automatically recording telephone calls he makes and television programmes he watches. It copies every web page he visits, transcribes every instant message he sends or receives, records the files he opens, the songs he plays, the searches he performs, the windows which are in the foreground of his computer at any time and the movements of the mouse, and uploads his position from the Global Positioning System (GPS) tracker that he wears constantly (Bell & Gemmell 2007, 45).

Unsurprisingly he has a large archive, of about 150 gigabytes, which he has used for tracing people he wants or needs to contact, finding citations for his academic papers, providing his medical history for his doctor and finding material for an obituary of a friend. MyLifeBits has also been important for experimentally discovering what extra tools are important for commercialising lifelogging, including face recognition to help with annotating photographs, speech-to-text software for transcribing or searching telephone calls and cleverer search and retrieval to classify documents in advance (Bell & Gemmell 2007, 45).

Information and social networking

The MyLifeBits model of lifelogging is pretty centralised and labour-heavy. If lifelogging is to take off as a pastime or as a way of life, the MyLifeBits model needs tweaking. In this section we discuss some more general points about what lifelogging is likely to entail.

Queries

It is likely that the competitive advantage of lifelogging is that the lifelogger can cross-reference information to provide greater power for his or her information retrieval. So for instance a digital photograph will be timestamped. The date and time of the photograph can be cross-checked against the lifelogger's personal calendar to discover what he or she was scheduled to be doing at that particular time. If the lifelogger has kept a GPS tracker, then the exact position of the camera can also be located. This information can be used to annotate the photograph with metadata to help in searching for and retrieving the picture (Tuffield et al 2006a).

Queries that might be suggested include:

- How many people attended the same events as I did last Summer?
- How many hyperlinks do I have in my email correspondence that I have yet to visit?
- What document was I reading on the plane to the World Wide Web Conference of 2008? What did I print off immediately beforehand?

- What was the name of the support band I saw in the Cully Jazz Festival of 2008?
- Who sent me emails while I was writing my paper for WWW08 and listening to Metallica?
- What pictures did I take at WWW08?

These are all plausible queries which have two vital things in common. First, they require integrated search of two or more data stores. And second, although they are (in the right context) serious questions, finding the answer through associative query and search demands keeping information that is of no intrinsic interest at all. If storage space was limited, the information upon which these queries sit would be the first to be filtered out.

Information

It is clear from the examples given in the discussion of MyLifeBits that the range of information that could be gathered is almost limitless. But of course most lifeloggers will be unable to afford a personal assistant to scan their coffee mug logos. Instead, the information that is likely to be gathered will be relevant to the lifelogger's main interests, and/or virtually costless to gather. Information sources likely to be popular include:

- Email. Sent and received.

- Calendar entries. A commonly-used personal calendar application is iCal, which is designed to be shared with other calendars over the Internet.
- Geodata. Many lifeloggers carry around a GPS unit to log positional data, and trackers are being incorporated into more devices (e.g. smartphones such as the Apple iPhone and the Nokia N95). This wouldn't always be of value – for example, it is hard for such a unit to track movement within a building or between buildings in close proximity. Network gazetteers, which provide information about locations of wifi hotspots being used by laptops, could fill in some of the gaps.
- Music downloads, and listening information.
- Web browser information, including bookmarks, downloads and navigation history.
- File system information, including document access information, such as lists of PDFs and Word documents read.
- Exchangeable Image File Format (EXIF) data from photographs. This is metadata automatically attached to digital photographs recording details such as the date and time they were taken, the camera type and the camera settings (e.g. focal length, aperture or shutter speed). The format allows GPS data to be recorded, although at the time of writing few cameras have built-in GPS receivers.

- Community-generated information. This could include photos taken by colleagues, friends and associates, events they have been to, music they (admit to) listen(ing) to, and tags placed on Web 2.0 content.
- Biometric sensors. This may be for medical applications. Microsoft's SenseCam has been used with biometrics to get interesting results (Kelly & Jones 2007).

Information of this sort is relatively straightforward to gather, and likely to be of interest in itself (to the individual who generated it), or to help with associative searching. It will help answer the key questions of context mentioned above: *what*, *where*, *when* and *who*; e.g. the question of *who* was relevant to a particular event might be answered with the help of emails, community tags, online accounts of events and friends' GPS records. Figure 1 shows schematically how information from different sources can be used associatively. For example, information available from an address book links the concepts of people and places; a calendar links people and time; events link people, time, places and transfers of money, and so on.

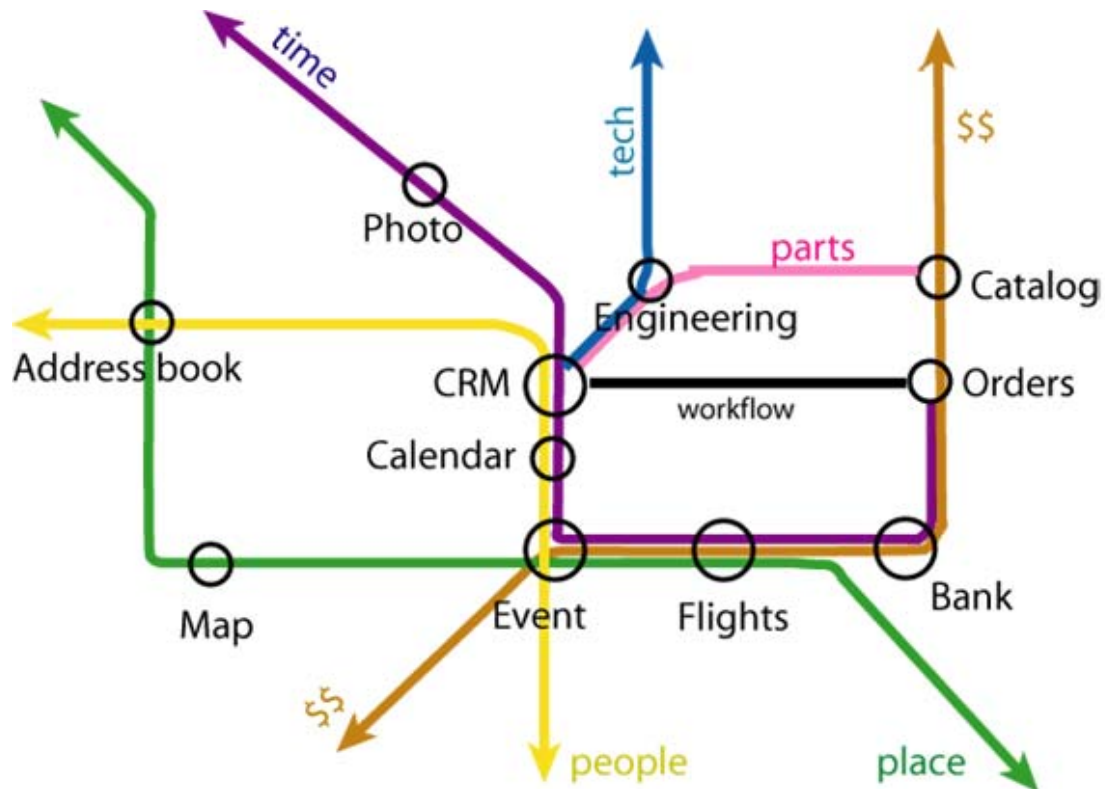


Figure 1: Integrated Data and Services (from Berners-Lee 2007)

Most lifelogging projects, such as MyLifeBits, tend to engineer over-arching knowledge representation formats to integrate information, but it is arguable that a simpler route is to retain the heterogeneity of information sources, so that applications using the information can use the most appropriate mappings between information sources, depending on which sources the application is currently exploiting (Tuffield et al 2006b).

Technologies

The requirement to query heterogeneous information implies that important underlying technologies for lifelogging will be those associated with the *Semantic Web* (Shadbolt et al 2006). One recent effort to develop low-effort lifelogging tools, the Semantic Logger, was aimed explicitly at using as many World Wide Web

Consortium Semantic Web recommendations as possible (Tuffield et al 2006b). The Semantic Logger is able to exploit Semantic Web formalisms as a *lingua franca* for representing information from large-scale, distributed and heterogeneous sources, which is the ultimate purpose of the Semantic Web. Such formalisms include the knowledge representation language RDF, querying language SPARQL, the framework of Universal Resource Identifiers (URIs) and basic structuring of information using the Friend of a Friend (FOAF) ontology (Tuffield et al 2006b).

Services

The rationale for collecting large quantities of generally unfiltered information is that it is cannot be specified in advance which information will be useful and which not, or what tasks the information might be used for. Hence the ideal architecture for a lifelogging system should be open not only from the perspective of information sources, but also from that of service provision. The Semantic Logger uses a knowledge representation store (using RDF) as a persistent repository for the system, and to mediate interactions between information sources and service outputs (Tuffield et al 2006b). On such a model, a software service could be devised for almost any purpose to which the stored information could be put. Likely examples include the following.

- Simple querying over the integrated data set, with questions such as those set out in the ‘Queries’ section above, which is the approach explored by MyLifeBits.

- Recommendation, where past behaviour is used to suggest items of interest. One's music downloads could suggest other music one might enjoy; academic papers saved to disc could suggest other items in the literature; the ratings one has given films could be used to suggest future films to watch or DVDs to buy (Tuffield et al 2006b). Relevant information might come from a wider social group as well as the lifelogger.
- The provision of metadata about information in order to facilitate search and retrieval (either by oneself or others), a strategy particularly suited to multimedia. One example, Photocopain, helps with the traditionally labour-heavy task of annotation of digital photographs (Tuffield et al 2006a).
- Creating and populating some kind of avatar, knowledgeable about oneself, which could act as an interface between oneself and other online actors, making decisions (for instance related to privacy and the revelation of information) on one's behalf. Such an avatar would be an interesting facet of one's identity that could, for instance, be used to construct narratives about one's life (which could easily take place after one was dead), to populate social networking sites, or partially to automate one's interactions with governments or companies (Wilks 2006).
- The provision of medical history. Possibilities include: (a) bodily sensor data to monitor changes in current health of the individual; (b) community-wide effort to prevent or monitor an epidemic; (c) the use of technology accurately to determine, e.g. someone's actual diet, a notoriously difficult thing to measure; (d) monitoring the use of household gadgets (e.g. kettles or fridges)

to signal that the user (perhaps an elderly person) remains in good health (Shadbolt & O'Hara 2008, 15-16). Google has a service to allow people to upload medical records (Krasner 2008), to follow Microsoft's HealthVault.⁸

- An aid to forgetting. Ironically, for a technology often seen as the antithesis of forgetting, lifelogging can be used to measure which pieces of information are recalled directly or used indirectly in associative recall. As well as being of interest from the information management point of view, this information could also be used to decide which information could be junked with least harm.

Social networking

The original rationale for lifelogging was as a personal tool to manage one's own information. However, since early experiments in this field, the practice of *social networking* has developed, where users generate and share information with others. This information can be quite specific in type (del.icio.us allows people to share their Web bookmarks) or form (Flickr allows sharing of photographs), or can be quite general (Facebook and MySpace allow people to connect and interact, revealing as much information about themselves as they care to post). Meanwhile, other practices such as blogging admit conversation, information and discussion into the public space.

In this context, information gathered by lifelogging practices could be shared, or enhanced by integration or cross-reference with information from others. As the

⁸ <http://www.healthvault.com/>.

recreation of context enriches information, there is no reason in principle why the information sources upon which such context-recreation draws should be restricted to ones controlled by oneself. Someone else's calendar might be as informative as one's own when it comes to retracing events in one's life.

This suggests the imperative to integrate the information describing people and their social relationships which is exposed by social networking sites. Portability of data across applications is already an important concern (Szomszor et al 2008) as people wish to carry data and personal profiles or identities across sites. Much of the information is non-sensitive and its creators are keen to share it – for instance, about film ratings and music downloads.

Hence it is probable, given the current profile of those who spend a significant proportion of their lives online, that the activities of lifelogging and social networking will intersect, and some social networkers will use lifelogging techniques to generate large quantities of information for their own use, and who will not be shy about sharing it with friends or like-minded people.

Privacy, empowerment and identity

Privacy concerns

Privacy is of course a serious issue for the individual who wishes to amass data. The course of DARPA's LifeLog project is instructive here. LifeLog was conceived as an experiment in life-long information capture, a fairly mainstream lifelogging effort: "an ontology-based (sub)system that captures, stores, and makes accessible the flow of one person's experience in and interactions with the world in order to support a

broad spectrum of associates/assistants and other system capabilities. The objective of this 'LifeLog' concept was to be able to trace the 'threads' of an individual's life in terms of events, states, and relationships" by aggregating raw data into a timeline that is an "episodic memory" (IPTO 2003).

Patterns of events in the timeline support the identification of routines, relationships, and habits. Preferences, plans, goals, and other markers of intentionality are at the highest level. (IPTO 2003)

So far, so harmless. But the solicitation for bids from a defence research agency was seen as worrying and many focused on the aim to build a database of all the transactions, including credit card details and phone calls, of an individual. DARPA had recently been under fire for its controversial Total Information Awareness system, which licensed too much surveillance for American voters even in the name of security, (O'Hara & Shadbolt 2008, 39), and the FutureMap system which appeared to encourage people to bet on the likelihood of terror attacks. LifeLog was pulled in 2004 under pressure from civil libertarians (Shachtman 2004).

LifeLog's difficulties are indicative of the strength of privacy fears when the funding body is not trusted by a potential user community. Anita Allen has reviewed a number of privacy issues that are likely to result from widespread lifelogging (Allen 2008), but at least some of the worries are overblown. First, there may be many occasions when the preservation of 'memories' is not what is required. Misfortunes or misjudgements would be preserved, possibly at the expense of average behaviour, creating in effect a false picture of a time period. But this is a danger of any kind of record – photographs of one's youth tend to cluster around parties and set-piece

meetings with other family members, neither of which activity actually looms very large at all on a day-to-day level. And the release of some sensitive information about the past – e.g. criminal records – is unlikely to be an issue with lifelogging. Practical obscurity is on the way out (O’Hara & Shadbolt 2008, 81-101), and lifelogging *per se* is not the most guilty party.

Secondly, Allen highlights complex issues to do with mental health and trauma, and worries that lifelogging might encourage pathological rumination about the past. It may be hard to persuade a patient that “lifelogger *capta* are not fixed, ‘hard’ evidence of an important whole story, rather than ... something partial, ambiguous, unimportant and interpretable”. As Allen points out, pathological rumination by those obsessed with the past can happen in the absence of any reliable memory at all, but one can go further and argue that the sheer wealth of detail of lifelogging might actually support therapy that stresses ambiguity. It is harder to maintain that one embarrassing photo tells the whole story of a gathering when another hundred are easily available. Similarly, Allen’s point that forgiving may be of limited value “when there is a diminished capacity to forget” cuts both ways. Forgiveness can be easier when there is a specific commitment to remember; reconciliation has been possible in Germany and South Africa partly because of a commemoration of the horrors of the holocaust and apartheid (Arendt 1958, Margalit 2002, Krog 1998).

Allen’s third worry is pernicious surveillance, and this of course is an important privacy issue. There are three routes by which lifelogging might become surveillance. First, lifelog data may feature the actions of others, who might appear in photographs, telephone calls, email exchanges and so on. Second, the tools for gathering data about oneself might also be the tools for gathering data about others. Third, governments

have a lot of power to insist that information that exists is made available to them; as Allen points out, current laws give the US government “access to virtually all means of communications and data storage” and there is no reason to believe that it would stop at lifelogs, or that it could be designed out of the technology (Allen 2008). These routes deserve serious consideration (though this is beyond the scope of this paper).

The empowerment of the individual

The privacy argument is clearly real. There is, however, also a flip side of such an argument. While it is clear that some people will get pleasure from storing such data, and that others will value improvements in data management, there are also genuine goods that come from constructing and maintaining an identity using lifelogging techniques. First, one has access to representations of the past which may come in useful. Of course, in any legal situation data would need to be authenticated, but one could prove one’s whereabouts if need be. An art professor at Rutgers, Hasan Elahi, who was arrested and subjected to some heavy treatment by the FBI in 2002 despite being absolutely innocent of any criminal or terrorist activity, has taken to lifelogging and posting the information on the Web as a pre-emptive alibi (Coughlin 2007).⁹

Second, the technology is of course helpful for the practice of what has been called ‘sousveillance’, community-based recording of events to democratise the process of surveillance. Rather than traditionally owned and controlled surveillance techniques being used to monitor a community, sousveillance supports the monitoring of the authorities, for instance searching for and reporting misdeeds by police forces, or electoral fraud, in a distributed way *by* a community. There are pros and cons to

⁹ <http://trackingtransience.net/>.

sousveillance (cf. Mann & Niedzviecki 2001, Brin 1998, O'Hara & Shadbolt 2008, 181-183), but for an individual it can be empowering to recall interactions whose nature is disputed.

Third, the construction and maintenance of an identity by an individual could act as a counterpoint to initiatives by formal and informal authorities to impose identities. There are many sources of unwanted identities, whether or not it is the creation of a formal system of ID cards (which may trespass on sensitivities, for example, by insisting on the use of a given name rather than the name in common use, or by failing to respect gender images of transgendered people), or an informal family insistence that one conform to social norms with respect to dress or sexual behaviour. The lifelog, for the lifelogger, might constitute the “real” person.

The Semantic Logger system has created a focus for the digital identity by the adoption of a Uniform Resource Identifier (URI) to refer to the user. Setting up a log requires the user to create a URI which will be associated with all the personal information logged. A FOAF document, developed using the FOAF ontology to model users and social networks, is imported into the personal knowledge base (KB) of the user, in which the primary subject of the document is the user's URI. The FOAF document can then point to other pieces of information about the user or his or her friends, so creating a linked structure of information available on the Web.

The result is an amalgamation of data about the user from distributed online sources in a single KB, providing the user with a global view of the personal information published on the Web. Information about the user can be gathered and associated, though not necessarily stored in one place, as separation is useful to help identify the

provenance of a particular statement. Although of course the published information is in the public domain, users can see the information collected, and can make informed decisions about whether to attempt to withdraw or amend items (Loizou et al 2008). Commercial systems in this space are beginning to appear, such as Garlik's QDOS.¹⁰

Logging in public and private

There are clearly complex balances to be struck. But there is a clear distinction between information about one that is out there, and personal information. Personal items, such as Web browsing habits, geodata or emails, can be stored in a personal knowledge base, while public domain information, including personal information that the user has deliberately exposed on social networking sites, can be stored separately. Although lifelogging involves gathering data in a relatively non-discriminating manner, that does not mean that one cannot discriminate in the treatment of the data. In particular, the distinction between publicly available data and data one does not wish to expose becomes more pressing as 'techy' lifestyles such as lifelogging and social networking are pursued simultaneously.

The Semantic Logger uses a three-way system to produce a very basic implementation of an intuitive trust model. It consists of a central KB where public data are held and published, while two password protected KBs are created for each user. The first holds the user's private data, while access to the second can be granted to friends. Hence – as with life offline – one's online identity as determined by the lifelog can be decomposed into entirely personal aspects, aspects reserved for one's intimates, and a public face (Loizou et al 2008).

¹⁰ <http://qdos.com/apps>.

Discussion

As lifelogging tools are currently a matter for scientific research rather than commercial application, there is a possibly brief window for reflection about privacy issues. Certainly the activity is rare enough that legislation is hardly required now. There have been some unrealistic suggestions, such as that of Dodge and Kitchin to program imperfections into the system, deliberate error that would prevent the lifelog from being veridical, and therefore invading privacy (Dodge & Kitchin 2007), and some possibly unfounded optimism, as with Cheng and colleagues, who are sanguine about the use of lifelog data by the judicial system (Cheng et al 2004).

There are certain ethical parameters that suggest themselves, as listed by Anita Allen.

No one should be required to keep a lifelog. No one should be suspected for not keeping a lifelog. Personal lifelogs should be deemed the property of the person or persons who create them. No one should record or photograph others for a lifelog without consent of the person or their legal guardian. A counter-technology to block lifelog surveillance should be designed and marketed along with lifeloggers. The owner/subject of a lifelog should be able to delete or add content at will. No one should copy a lifelog or transfer a lifelog to a third party without the consent of its owner. (Allen 2008)

Tools to gather data will have obvious applications for surveillance, and this is a powerful worry, as evinced by the fate of the DARPA LifeLog project. It is essential that one is not forced (either by law or social pressure) to keep tabs on oneself. But discussions about privacy and lifelogging have made unwarranted assumptions which

skew the privacy arguments. Many presume that the data is necessarily personal, though it may be public domain. And why think that information gathered should be kept private? The publication of the information might just be the point.

There are also practical considerations – the appearance of others in our photographs is not something we can easily control. And it is important not to develop solutions that apply to outdated technologies. For instance, large amounts of an individual's information are migrating from their personal standalone devices, and are being stored instead in what is becoming known as 'the cloud' – a set of computing resources operated by a third-party provider (such as Google) and located in data centres, while the user is unconcerned with the actual technology being used, and instead hires the information storage or processing service.

In this more nuanced environment, lifelogging tools can provide the user with somewhat more control, gathering together his or her appearances in the public domain. Knowing what can be seen is an important first step in the preservation of privacy, and it is important that restrictions or distortions are not brought into lifelogging tools that prevent this empowerment of the user. One's online identity has public and private aspects, and focus on the private aspects will be misleadingly partial. Furthermore, the social aspects of the Web are where its interest lies for many people. Restrictions on their ability to construct their own identity would be a severe curtailment of their personal freedom.

A final point: lifelogging sounds like a somewhat recondite pursuit, but personal knowledge management is an issue for anyone with a significant Web presence, or who uses digital technologies. Lifelogging is an extreme case, but the tools and

interfaces that support it will also support querying of and retrieval from smaller repositories of personal data collected using more discriminating methods, a fact that is captured by the term ‘Personal Information Management’. To that extent, lifelogging tools are tools for everyone to exert more control over their personal data, their public presence online and their digital identity.

Abbreviations

DARPA: Defense and Advanced Research Projects Agency

EXIF: Exchangeable Image File Format

FBI: Federal Bureau of Investigation

FOAF: Friend of a Friend

GPS: Global Positioning System

KB: Knowledge Base

M4L: Memories for Life

PDA: Personal Digital Assistant

PDF: Portable Document Format

PIM: Personal Information Management

RDF: Resource Description Framework

SPARQL: SPARQL Protocol and RDF Query Language (sic: a recursive acronym)

UKCRC: United Kingdom Computing Research Committee

URI: Universal Resource Identifier

References

M. Ahmed, H.H. Hoang, M.S. Karim, S. Khusro, M. Lanzenberger, K. Latif, E. Michlmayr, K. Mustofa, H.T. Nguyen, A. Rauber, A. Schatten, M.N. Tho & A.M. Tjoa (2004). 'SemanticLIFE: a framework for managing information of a human lifetime', *6th International Conference on Information Integration and Web-Based Applications and Services (IIWAS)*, Jakarta, Indonesia, <http://storm.ifs.tuwien.ac.at/publications/iiwas2004.pdf>.

Anita L. Allen (2008). 'Dredging up the past: lifelogging, memory and surveillance', *University of Chicago Law Review*, 75, 47-74.

Hannah Arendt (1958). *The Human Condition*, Chicago: University of Chicago Press.

Gordon Bell & Jim Gemmell (2007). 'A digital life', *Scientific American*, March 2007, 40-47.

Tim Berners-Lee (2007). *Looking Back, Looking Forward: The Process of Designing Things in a Very Large Space*, Inaugural lecture, Southampton University, <http://www.w3.org/2007/Talks/0314-soton-tbl/>.

Emma Berry, Narinder Kapur, Lyndsay Williams, Steve Hodges, Peter Watson, Gavin Smyth, James Srinivasan, Reg Smith, Barbara Wilson & Ken Wood (2007). '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, 582-601.

David Brin (1998). *The Transparent Society: Will Technology Force Us To Choose Between Privacy and Freedom?* New York: Basic Books.

William C. Cheng, Leana Golubchik & David G. Kay (2004). 'Total Recall: are privacy changes inevitable?' in *Proceedings of the Capture, Archive and Retrieval of Personal Experiences Workshop (CARPE) at ACM Multimedia 2004*, New York, <http://bourbon.usc.edu/iml/recall/papers/carpe2k4-pub.pdf>.

Kevin Coughlin (2007). 'Tracking himself, so the FBI won't have to', *Digital Life with the Star Ledger*, 28th October 2007, http://blog.nj.com/digitallife/2007/10/tracking_himself_so_the_fbi_wo.html.

Alan Dix (2002). 'The ultimate interface and the sums of life?' *Interfaces*, 50, 16.

Martin Dodge & Rob Kitchin (2007). 'Outlines of a world coming into existence: pervasive computing and the ethics of forgetting', *Environment and Planning B: Planning and Design*, 34, 431-445.

IPTO (2003). *LifeLog Proposer Information Pamphlet*, Defense Advanced Research Projects Agency Information Processing Technology Office document PIP_03-30, http://web.archive.org/web/20030603173339/http%3a//www.darpa.mil/ipto/Solicitations/PIP_03-30.html.¹¹

Krissi M. Jimroglou (1999). 'A camera with a view: JenniCam, visual representation and cyborg subjectivity', *Information, Communication and Society*, 2, 439-453.

William Jones (2008). *Keeping Found Things Found: The Study and Practice of Personal Information Management*, Burlington MA: Morgan Kaufmann.

Vaiva Kalnikaitė & Steve Whittaker (2007). 'Software or wetware? Discovering when and why people use digital prosthetic memory', *Proceedings of Conference on Human Factors in Computing Systems (CHI07)*, New York: ACM Press, 71-80.

Liadh Kelly & Gareth J.F. Jones (2007). 'Venturing into the labyrinth: the Information Retrieval challenge of Human Digital Memories', *Workshop on Supporting Human Memory With Interactive Systems*, at *The 21st British Human-Computer Interaction Group Annual Conference*, Lancaster, UK.
<http://diuf.unifr.ch/people/lalanned/MeMos07/files/kelly.pdf>.

¹¹ This is an archived version of the document. The official version of the document has been taken offline by DARPA.

Jeffrey Krasner (2008). 'Google unveils medical records storage plan', *The Boston Globe*, 20th May, 2008.

http://www.boston.com/business/healthcare/articles/2008/05/20/google_unveils_medical_records_storage_plan/.

Antjie Krog (1998). *Country of My Skull*, Parktown, South Africa: Random House.

Antonis Loizou, Mischa M. Tuffield, David P. Dupplaw, Paul H. Lewis, Nigel R. Shadbolt & Srinandan Dasmahapatra (2008). 'Context and content in multimedia knowledge management: an OpenKnowledge approach', submitted to *IEEE Transactions on Multimedia*.

Steve Mann & Hal Niedzviecki (2001). *Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer*, New York: Random House.

Avram Margalit (2002). *The Ethics of Memory*, Cambridge MA: Harvard University Press.

Kieron O'Hara, Richard Morris, Nigel Shadbolt, Graham J. Hitch, Wendy Hall & Neil Beagrie (2006). 'Memories for Life: a review of the science and technology', *Journal of the Royal Society Interface*, 3, 351-365.

Kieron O'Hara & Nigel Shadbolt (2008). *The Spy in the Coffee Machine: The End of Privacy As We Know It*, Oxford: Oneworld.

Daniel L. Schacter (2001). *The Seven Sins of Memory: How the Mind Forgets and Remembers*, New York: Houghton Mifflin.

Noah Shachtman (2004). 'Pentagon kills LifeLog project', *Wired*, 2nd April, 2004, <http://www.wired.com/politics/security/news/2004/02/62158>.

Nigel Shadbolt, Wendy Hall & Tim Berners-Lee (2006). 'The Semantic Web revisited', *IEEE Intelligent Systems*, 21(3), 96-101.

Martin Szomszor, Harith Alani, Ivan Cantador, Kieron O'Hara & Nigel Shadbolt (2008). 'Semantic modelling of user interests based on cross-folksonomy analysis', submitted to *International Semantic Web Conference 2008*.

Mischa M. Tuffield, Stephen Harris, David P. Dupplaw, Ajay Chakravarthy, Christopher Brewster, Nicholas Gibbins, Kieron O'Hara, Fabio Ciravegna, Derek Sleeman, Nigel R. Shadbolt & Yorick Wilks (2006a). 'Image annotation with Photocopain', *Proceedings of the WWW06 Workshop in Semantic Web Annotations for Multimedia (SWAMM '06)*, Edinburgh, UK.

Mischa M. Tuffield, Antonis Loizou, David Dupplaw, Srinandan Dasmahapatra, Paul H. Lewis, David E. Millard & Nigel R. Shadbolt (2006b). 'The Semantic Logger: supporting service building from personal context', *Proceedings of the 2nd Capture, Archive and Retrieval of Personal Experiences Workshop (CARPE2006)* at *ACM Multimedia 2006*, Santa Barbara, USA.

Yorick Wilks (2006). *Artificial Companions as a New Kind of Interface to the Future Internet*, Oxford Internet Institute Research Report 13,

<http://www.oii.ox.ac.uk/research/publications/RR13.pdf>.