Lifelogging: Privacy and Empowerment with Memories for Life

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

Version of article appearing in *Identity in the Information Society*, 1(2)

http://dx.doi.org/10.1007/s12394-009-0008-4

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 2582

Abstract: The growth of information acquisition, storage and retrieval capacity has led to the development of the practice of *lifelogging*, the undiscriminating 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 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, M4L, Personal Information Management, privacy, identity, surveillance, social networking

Introduction

Technology has always been important as a memory aid but 21st century computing power has provided the means for augmenting memory to a degree hardly dreamt of even ten years ago. Information can be stored in enormous quantities, so people no longer have to be selective. Technology has nearly reached the stage when all information, interesting or otherwise, generated in a lifetime by a single person can be assembled and queried relatively efficiently, creating a need for *Personal Information Management* (PIM – W. Jones 2008). This is 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 (O'Hara et al 2006).[1]

The logical end-point of the drive to store digital information is to store indiscriminately, a practice called *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.

[1] http://www.bcs.org/server.php?show=ConWebDoc.5203.

The range of information that can be gathered is almost limitless, but in practice the information that is *likely* to be gathered will be relevant to the lifelogger's main interests, and/or very cheap to gather. Information sources likely to be popular include email both sent and received, electronic calendar entries, music downloads and listening habits, Web browser information, including bookmarks, navigation history and downloads, file system information including document access, EXIF data from photographs and biometric sensors, which may be used during medical treatment. It is likely that community generated information will become increasingly important. This could include photos taken by colleagues, friends and associates, events they have been to, music they listen to, and Web 2.0 style tags placed on content.

One more important type of data is geodata. Some 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. An example application which is making use of a gazetteer of wifis and their physical locations to expose geo-information is Loki's Skyhook[2], currently used in the Apple iPhone, which implements the W3C Geolocation API[3] to fill in the aforementioned gaps. Compass directions of photographs are also of interest here, as GPS data tell you where the camera is, not what has been photographed, but a direction plus an indication of focus could identify the surroundings of a photograph more accurately (Naaman et al 2004). Finally, gazetteers of places are very helpful for

[2] http://loki.com/blog/.
[3] http://dev.w3.org/geo/api/spec-source.html.

searching, because a mapping from GPS coordinates to actual place names is more meaningful; nothing about a pair of GPS locations tells you that both are in London, for instance.

The purpose of storage of such information can vary, and may not be clear even to the lifelogger at the point of storage, but it is relatively straightforward to gather and likely to be of interest either 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 shared location-based information.

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 can facilitate 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 lifelogging, and discuss the concerns about privately held 'Memories for Life'. The next section will profile some of the more prominent efforts in this space, as well as setting out some principles and technologies. Then we shall consider some of the social aspects of lifelogging. Next, we shall examine some of the issues surrounding privacy, identity and accountability for the lifelogger. Finally, we

will end with a brief discussion of the perceived problems with lifelogging technology, and a short conclusion.

Lifelogging

Lifelogging is not new. One important effort was LifeLog, sponsored by the American Defense and Advanced Research Projects Agency (DARPA), conceived as an experiment in life-long information capture. Its governing principle was "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). The involvement of a defence research agency was seen as worrying and critics focused on the aim of building a database of all the transactions, including credit card details and phone calls, of an individual. LifeLog was cancelled in 2004 under pressure from civil libertarians (Shachtman 2004), although military planners and technologists remain alive to the possibility of harnessing information gathered in routine manner, as for example with the slogan 'every soldier a sensor' (Magnuson 2007).

The most famous and comprehensive experiment is the attempt by Gordon Bell of Microsoft to create a digital archive, MyLifeBits.[4] This began in 2001, but its roots were in Bell's attempt from 1998 onwards to digitise all documents in all media, a high-cost strategy that required a team of full-time personal assistants. MyLifeBits was then set up to enable Bell to make sense of the enormous repository of information with querying, search, retrieval and new capture tools (Bell & Gemmell 2007). Steve Mann uses wearable computing to record his daily existence, and has

[4] http://research.microsoft.com/barc/mediapresence/MyLifeBits.aspx.

enjoyed media status as 'the world's first cyborg' (Mann & Niedzviecki 2001).[5] His latest efforts, using the EyeTap visual memory prosthetic, in effect turning his eye into a camera and his body into a Web server, he calls 'cyborglogging' or 'glogging'.[6] Jennifer Ringley also achieved celebrity and indeed notoriety when she set up JenniCam, a webcam that recorded events in her living space, which 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 private moments, but for most of the experiment, which ran from 1996-2003 the images were unfiltered.[7] Other initiatives include Total Recall[8] (Cheng et al 2004) and SemanticLIFE[9] (Ahmed et al 2004).

Semantic Web technologies

Most lifelogging projects, such as MyLifeBits, tend to engineer overarching 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 sources, depending on which the application is currently exploiting (Tuffield et al 2006b).

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

[5] http://wearcam.org/.
[6] http://wearcam.org/glogs.htm.
[7] http://www.arttech.ab.ca/pbrown/jenni/jenni.html.
[8] http://bourbon.usc.edu/iml/recall/.
[9] http://storm.ifs.tuwien.ac.at/.

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).

Adopting Semantic Web technology will allow for lifelogs to be contextualised and grounded with respect to web accessible linked data,[10] by making use of resources such as dbpedia,[11] a machine-readable version of Wikipedia,[12] to give shared identifiers to places one has visited, or MusicBrainz,[13] an open content music database, for identifiers for the music they have listened to. Such grounding of terms promotes sharing and linking of information resources across communities.

Services

The rationale for collecting large quantities of generally unfiltered information is that it cannot be specified in advance what information will be useful or 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

[10] For a diagrammatic vision of the 'cloud' of linked open data, see http://richard.cyganiak.de/2007/10/lod/.
[11] http://dbpedia.org/.
[12] http://www.wikipedia.org/.
[13] http://musicbrainz.org/.

model, software services could be devised to exploit the information, as well as simply querying over the integrated data set.

One such service is 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 from the lifelogger. Many social sites already exploit information for recommendations, such as Amazon and last.fm, but they do not take other information into consideration. Recommendations can be finer-grained than is allowed for by such sites, because preferences change with location, circumstances, etc. For instance, if someone listened to jazz at home, but not in the car, where they preferred louder rock music, a more sensitive system could spot the pattern and recommend downloads accordingly.

Another service would be to provide 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).

A third possibility would be 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 revealing 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, to

populate social networking sites, or partially to automate one's interactions with governments or companies (Wilks 2006).

The provision of medical history is a fourth example. Possibilities include: (a) sensor data to monitor changes in current health of the individual, or so-called telemedicine; (b) community-wide effort to prevent or monitor an epidemic; (c) the use of technology accurately to determine someone's 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 a frail person) remains in good health (O'Hara & Shadbolt 2008, 15-16). There are services which allow people to upload medical records, such as Google Health[14] or Microsoft's HealthVault,[15] while social networking methods have been used to support Web 2.0-style information sharing about symptoms and treatments of various illnesses, such as PatientsLikeMe,[16] which hosts a number of MySpace-like communities discussing conditions such as Motor Neuron Disease or Parkinson's.

Non-solipsistic lifelogging

The MyLifeBits model is centralised and labour-heavy, and needs tweaking if lifelogging is to take off as a pastime or as a way of life. In general, it has been often assumed that lifelogging is a solipsistic activity, hoarding information about oneself for one's own purposes. This, we argue, is likely to be a mistaken assumption, even if it is true of many (but not all) of the early pioneers of the practice.

[14] http://www.google.com/health.
[15] http://www.healthvault.com/.
[16] http://www.patientslikeme.com/.

The original rationale for lifelogging was as a personal tool to manage one's own information. However, since the early experiments the practice of *social networking* has developed, where users generate and share information with others. This information can be quite specific in type, such as del.icio.us which allows people to share their Web bookmarks. The site can be based around a specific medium, such as YouTube which allows sharing of video, or Flickr which allows sharing of photographs. Or the networking site can be quite general, such as Facebook and MySpace which 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 re-creation of context enriches information, there is no reason in principle why the information sources upon which 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 there is mileage in integrating 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 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, film ratings and music downloads.

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. Some social networkers will use lifelogging techniques to generate large quantities of information for their own use and to share with friends or like-minded people, while also incorporating social information to add further contextual information to their own lifelog.

There are complex balances to be struck in this area, 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 rule out discrimination 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.

In such an environment, privacy protection is clearly an imperative. The Semantic Logger uses a three-way system to produce a 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 (Tuffield 2006b). More complex systems, for example exploiting FOAF, could

implement restrictions, for instance, to people who are friends or friends of friends but no further down the chain.

Social lifelogging developing extends the solipsistic model by adding in shared information, and sharing information one generates oneself. This augmented lifelog will be harder to control. It will be possible to delete information over which one has control (one's own photographs or blog), although if others are linking to it that may become bad manners. If one's lifelog links to someone else's information, then that cannot necessarily be deleted, although one could remove links to it. The picture of oneself that emerges will be more socially-based, and less inwardly-directed.

Furthermore the information will contain a number of important organising principles, rather than simply being an unstructured dataset. The use of Semantic technologies will support the exploitation of metadata mediated by ontologies. The metadata may take the form of personal metadata or social metadata, depending on whether it is captured and stored directly by the lifelogger, or harvested from the Web.

As an example of the use of social metadata, consider the exploitation of geolocation metadata by the Semantic Logger (Tuffield 2006b). Location information is taken from the photo-sharing site Flickr, in the form of geo–tagged images, and from the social networking site Facebook. Through the use of linking events, based on iCal and Facebook events, it is possible to suggest that two Facebook friends may have been at the same location. This piece of social information would allow for images taken at the same time to share tags, and hence increase the available contextual information surrounding that piece of multimedia.

Privacy, identity and accountability

Privacy concerns

It is of course hard to predict precisely what the trajectory of such online social developments will be. The assumption of a series of concentric ‘privacies’ in this space may also be too strong, though the simplicity of the Semantic Logger’s model is attractive and not demanding. Privacy tends not to be concentric, but generally takes the form of something like a Venn diagram of interlocking and overlapping spaces, where the Semantic Logger’s transitivity assumption (if X can have access to p, and Y is a closer friend than X, then Y can have access to p) does not hold. Indeed, there is empirical evidence from the blogging world that quite often diarists actively prefer strangers to read their material, but discourage intimates (Sorapure 2003).

The rise of social networking may well have downgraded the instrumentality of trust. The general problem of trust is to connect, preferably causally, trust and trustworthiness. A causal connection can go in two directions: one is from trustworthiness to trust, which we might call *Weberian*, and the other is from trust to trustworthiness, which we might call *Durkheimian*. The Weberian method is the default for the Web and Semantic Web (in which trust plays a prominent part – Shadbolt et al 2006), where the would-be trustee begins by being trustworthy, and to be trusted must prove trustworthiness via some mechanism, which might be reputation, credentials or whatever (Golbeck 2006). The idea is to minimise risk, but risking opportunity cost of a failure to trust.

However, the Durkheimian model, where one trusts another, thereby welcoming him or her into one’s moral community, which then generates (as people are social beings)

trustworthy behaviour, is becoming more popular particularly in the world of social networking. People put out social information on the possibly implicit assumption that although risks of harm are present, the dangers of harm are small. A few people will engage deeply with one's personal information, and will therefore be sympathetic. Most people will be uninterested or positively hostile, but, life being short and their being strangers, they will merely ignore one's website, rather than actively seek to trash it, vandalise it, leave offensive comments or misuse information revealed (de Laat 2008). A common variation on this is the thought that malign behaviour is actually more likely to be shown by friends (possibly offended by comments or intimate thoughts) and offline acquaintances, rather than strangers (Sorapure 2003).

The social networking world seems to be governed by Durkheimian rules different from more instrumental views of the Web which tend to follow the Weberian model (although whether this is driven by Web designers rather than Web users is perhaps a moot point). As Durkheimian trust is risky, we cannot rule out the possibility that social networking is going through a particular 'moment' when it appears to be sustainable, and that attitudes and perceptions will change and a more suspicious culture arise.

It is clear, particularly on the reasonable conceptualisation of privacy as a condition of restricted access to information about a person (Allen 1988), that the Durkheimian trust that is exhibited in social networking generally, and social lifelogging in particular, will threaten privacy. Anita Allen has reviewed a number of privacy issues that are raised by lifelogging (Allen 2008), and though the issues she raises are certainly real, her examples of lifelogging are based on the existing literature, rather than the likely trajectory of lifelogging practice. In the socio-technical context we

have stressed, where lifelogging is not necessarily a solipsistic activity, and information coverage is partial and opportunistic rather than dedicated and maximised, the worries appear overblown to some extent.

First, she points out, there may be many occasions when the preservation of ‘memories’ is a bad idea. Misfortunes or misjudgements would be preserved, creating in effect a false picture of a time period. It is true that unwanted as well as wanted information is likely to be preserved, but the situation with lifelogging may be less pernicious than other methods of preserving the past. A traditional collection of photographs will also give a skewed version of past events, but with the added problem of a lack of context.

Allen gives the example of someone slapping a friend at a party: the incident is captured by a dozen bystanders and the story leaks out. If that person deletes the incident from her lifelog, she will still be prevented from forgetting the incident, even if she has been forgiven. This is a regrettable situation to be sure, but not one unique to lifelogging. In these days of digital photography, such circumstances are becoming common; surveillance has been democratised. Surely the lifelog has the advantage that, by being a *richer* account of the incident, it can at least be put into some sort of perspective. The slap can be placed in the context of the friendliness of the rest of the party, possibly the proximate cause (drink?), the forgiveness, the regret and the long years of friendship before and after the incident.

Allen is correct to say that “people … have a legitimate moral interest in distancing themselves from commonplace misfortunes and errors,” but it seems incorrect to add, as she does, that “in order to create that distance, they need to be safe from memory.”

Actually, the distancing process must involve coming to see that the incident was atypical, uncharacteristic and not to be repeated. It is a mistake from which a lesson should be learned, and for that, the incident needs to be seen as a whole – which cannot happen with the single snapshot, but might with a lifelog which draws information from public as well as private sources. Merely forgetting, on the other hand, seems to rule out having learned the lesson.

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". Again, one cannot deny the nugget of truth in this thought; equally, as Allen herself points out, pathological rumination by those obsessed with the past can happen in the absence of any reliable memory at all, never mind partial and ambiguous ones.

Furthermore, our reply to Allen's first worry also applies here. The sheer wealth of detail in a lifelog might actually support therapy. If the therapist wished to stress the ambiguity of the past and the partiality of the patient's memory of it, what better way than to provide a richer account? It is harder to maintain that one embarrassing photo tells the whole story of a gathering when another hundred unembarrassing ones are easily available. Different patients doubtless need different therapies, but it seems to us at least plausible that a richer story will be valuable on some occasions to contextualise a patient's pathological rumination.

Allen's third worry is pernicious surveillance, and we do not wish to deny that this is an important issue. There are three routes by which lifelogging might become surveillance. First, lifelog data may feature the actions of others in photographs, telephone calls, email exchanges and so on. Second, the tools for gathering data about oneself might also become 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).

A related issue is that lifelogging might be a private good that damages the public good by creating costs that do not accrue to the individual lifelogger. Bailey and Kerr (2007) have argued that the law relating to privacy has a connection with the 'reasonable expectations of privacy' within the relevant society. It may well be that the practice of lifelogging could decrease those reasonable expectations of privacy, and therefore undermine privacy protection.

There clearly is a danger that lifelogging, if it ceased to be a minority pursuit, could develop unpleasant overtones. Allen is correct that governments are likely to want to appropriate whatever information exists – they are not noted for restraint in this area – but there may be an argument to say that lifelogging technology in itself is neutral. Yes, people may incriminate themselves, but equally they may do so with any record they make of their activities. Yes, others may be incriminated, but equally they may be so by any kind of record. The lifelog, being a richer account, may be more incriminating, possibly, than other types of record (e.g. CCTV pictures), but equally might be less incriminating, providing essential context and removing circumstantial

ambiguities. Indeed, as well as providing a case for the prosecution, a lifelog might also provide a case for the defence. We will return to this point in more detail below. Yes, spying might be a problem, but by a parallel argument, if lifelogging became a widespread practice, then spies or obsessives might be easier to spot, because they would feature on others' lifelogs.

If expectations of privacy have fallen, then people are more likely to be on their guard – awareness is a sort of protection against surveillance. However, this does bring us to Bailey and Kerr's important point (2007) about the decline of reasonable expectations causing a related decline in legal privacy protection. They note that Jennifer Ringley originally wanted to present a realistic picture of the life of a young woman, to show that it was not like an episode of *Friends*. JenniCam certainly was neither as entertaining nor as false as that show, but Ringley could not keep control of her information, and soon pornographic images filtered out of the total set began appearing out of context. The total picture soon became the edited picture.

There is a question, which has recurred throughout this section, as to whether lifelogging makes the current situation, where surveillance is constantly increasing, any worse. At the moment, one might reasonably expect to appear on some record of events if one appears in public, and one would be wise to conduct oneself accordingly. Lifelogging may increase the probability that one actually did appear on a record, assuming the practice becomes widespread, but it is not clear that lifeloggers would impinge on a *different* class of situations.

One already should expect that one's emails, Web browsing history, Web 2.0 content and iCal entries will be visible for a long time. Appearing in public makes one liable

to appear in digital photographs already, as well as CCTV footage. Other types of lifelogger information, such as geodata, document access records and biometric sensors are unlikely to convey much extra about other people. There may be some lifelogging technologies, such as Microsoft's Sensecam pioneered by Bell, that could increase the probability that one's privacy was infringed, but currently these are not, and may never be, widely used. Hence, although it is certainly a problem that reasonable expectations of privacy are in decline, and Bailey and Kerr are right to draw attention to this, at the moment there seems no reason to think that lifelogging will make a serious contribution to it, compared say, to the intrusive surveillance activities of governments.

The dialectic between public and private is highly complex. It is arguable that Jennifer Ringley was not properly aware of the nature of the threat to her privacy posed by the JenniCam (Bailey & Kerr 2007), but a different model of lifelogging would focus on information both public and private, with two-way traffic between the individual and his or her social network(s). Although Ringley published information routinely and indiscriminately, that is not an essential part of lifelogging. Social networks and formalisms such as FOAF allow one to relinquish some control over one's information, while remaining within one's comfort zone – much as happens in offline life where several external identities can be maintained to interact with friends, family, lovers, enemies, workplace colleagues and so on. Ultimately it may be that lifelogging does not reduce reasonable expectations of privacy, but rather that reasonable expectations of privacy will dictate lifelogging practice.

If the days of practical obscurity are numbered by the development of digital technologies, then arguably it is preferable (a) to provide rich rather than sparse

accounts of an incident, and (b) to provide access to diverse sources of information, in order to allow corroboration or otherwise of particular pieces of data. In each case, lifelogging, all things being equal, is more likely to fit the bill.

Identity

The privacy argument is clearly real, but it must be offset against the empowerment of the individual that lifelogging can provide. Perhaps the most important way in which this can happen is to give the lifelogger sufficient control over his or her information to act as a counterpoint to initiatives by formal authorities – and informal ones, such as families, too – to impose artificial identities. There are many sources of unwanted identities, whether or not it is the creation of a formal system of ID cards, a financial identity 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.

Against this, it may be argued that it is hardly an *identity*. In the first place, the indiscriminateness of information collection undercuts claims that anything interesting has been *constructed*, while secondly it has yet to be established that the lifelog is anything more than a data set. This is certainly a valid point if a digital identity is defined as a structured set of parameter values specified in advance, either held on some database held to be canonical (as with an official ID) or that can be associated with a high degree of accuracy with some personal and unchanging characteristic (a biometric). In other words, a system of identification is set up in advance, and the subject surrenders some information in order to be identified by the system.

Such an approach can be valuable in conceptualising and defusing the privacy issue. We do not want to develop a full-scale informational identity theory here, but we would like to sketch a position that renders the claim that the lifelog is an identity plausible. To begin with, consider an argument made by Floridi that privacy issues emerge because of an apparent contradiction between a desire to maximise information flow, because of the benefits this brings, and a desire for a high level of information protection to preserve privacy (Floridi 2005, 197-198). His answer is to draw a distinction between the sharable information about oneself which can flow freely, and the ontic data that "constitute someone," a much smaller set that can and should be restricted without undermining the benefits of a high information flow. Floridi's ontological theory specifically considers each person as constituted by his or her information (Floridi 2005, 194), and argues that this is a way forward from unsatisfactory privacy theories that either seek to boost the individual's control over his/her information, or to restrict access to that information for everyone else.

However, the problem with this is that, if we accept the claim that "we are our information," the result is unsatisfactory when combined with Floridi's privacy argument. A person is a complex moral and physical being, who may have important properties that vary according to the people with whom he/she comes into contact, and will certainly have varying moral attitudes and obligations. The complexity of personhood is unlikely to be capturable either by a one-size-fits-all database or biometric – indeed, a quote from Proust used by Floridi (2005, 194-195) undercuts the notion of a constitutive biometric rather neatly. This is not to say that simple ID systems are not entirely adequate for restricted functions in particular contexts, only that the ontic claim depends on respecting those restrictions. If an identity is required

that functionally goes beyond the legitimation of e-voting or the use of a credit card online then a wider notion is surely required.

Such notions are available in the Lockean tradition that locates personal identity in consciousness, specifically memory. In this tradition, the lifelog has an important role to play in supporting the lifelogger's own identity, as it collects information that the lifelogger him- or herself has targeted to support memory. The lifelog need not be *the* identity, but provides a wide range of materials for the lifelogger to deploy and edit. Nonetheless, the lifelogger cannot be in total control; analogous to the way in which a Lockean identity can force someone to face up to responsibilities and less savoury aspects of the past, the lifelog is likely to contain reference to unfortunate aspects of the past that could be edited out but need not be. It may be too time-consuming to, e.g. remove all evidence of a drunken exploit, a prison sentence or a less-than-fondly-remembered boyfriend or girlfriend, and the use of publicly-held information as well will make it even more difficult.

The lifelogger has sufficient control to develop other personae that not only counter official or unwanted identities, but are powerful enough to achieve things in the online world. We do not suggest that this sketch constitutes a well-worked-out theory of personal identity, and indeed the idea that a lifelog constitutes, or contributes significantly to, identity is not crucial to the main claim of this paper, that many of the privacy concerns surrounding lifelogs are overblown.

It should also be pointed out that the identity need not be as diffuse as the picture suggested above, and could be focused around some highly specific data or identifier which would go some way to meet Floridi's suggestion to defuse the privacy problem.

The Semantic Logger system creates a focus for the digital identity by the adoption of a Uniform Resource Identifier (URI) to refer to the user, consistent with the use of Semantic Web technology. 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. This is in line with recent thinking about the creation of decentralised Web-compatible identity protocols, which can be harnessed to various security and trust-enhancing protocols (Weitzner 2007). In particular, the existence of the URI need not impinge on privacy; if we refer back to the example of the use of social geolocation metadata in the previous section, Flickr reflectors expose user-based information in HTTP-resolvable RDF, using the SOIC ontology, but people surfing a public Flickr page or a publicly available RDF representation of a Flickr page will not have enough metadata to connect the information with the owner's public FOAF URI, unless they already know it.

The Semantic Logger produces an amalgamation of data about the user from distributed online sources in a single KB, providing a global view of 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. The

monitoring of the Web to protect one's good name against what is becoming known as 'cyberspite' (e.g. posting negative comments on auction sites about people who have posted negative comments about you) has become an area of commercial interest (D. Jones 2008).

Accountability of the individual

There are other positive effects of lifelogging, and in the rest of this section we will briefly review two of them. First, as well as material to create identities to counter imposed identities, the lifelogger has material to create countervailing representations of the past. For example, an art professor at Rutgers, Hasan Elahi, who was arrested in 2002 on serious terrorism allegations 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).[17]

Accountability is an increasingly prominent aspect of life, and indeed Allen has argued strongly that it should be not only a fact, but a value (Allen 2003). She notes that the spread of social, economic and political freedoms combined with ambivalence about privacy has led to a culture in which it is often expected to make public apparently intimate details about one's life, now that one is more or less safe from state punishment. Freedom and openness of conduct results in more people knowing, and wanting to know, about it. One may deplore this increase in accountability, but as a trend it is hard to ignore. Even if one does ignore social pressure, there is still accountability in law, to courts, to the media, to one's intimates, to government, and if

[17] http://trackingtransience.net/.

Allen is correct, to those groups of people who share aspects of one's identity (e.g. one's race, one's gender – Allen 2003).

By accountability, Allen means the expectation or requirement to inform others of, explain or justify conduct; depending on the context, one may also be subject to punishment. In a situation where the lifelogger is accountable for something, he or she is empowered by the lifelog to the extent that it allows him or her to provide information to satisfy the expectation, or provide evidence that competing accounts are false. The information provided would have to be authenticated of course, but even so at least one has access to the basis of an account.

So, to take the example of Elahi's arrest, had he been lifelogging earlier, he would have been able to construct an account of his activities to counter the FBI's. In a fraught legal situation such as being questioned on terror charges, it is not enough to counter the authorities' theories. One needs a certain level of proof, and the stronger the evidence against one, the better one's proof needs to be. Indeed, a lifelog may well be exceedingly rich in detail – we have emphasised the potential for the amalgamation of information from many different sources – and so aspects of the lifelog may well corroborate others. The Flickr photos may support GPS data, which may support an iCal entry which may support the veracity of the content of blogs or emails. Furthermore, social lifelogging as we have described gives an element of objectivity to the account. In general, the richer the evidence base of an account, the more convincing it will tend to be.

The empowerment is not of course total. A lifelog will be of most use when the facts are in dispute, but sometimes it is the explanation of behaviour that is at issue. Lifelog

information tends not to reflect interior or cognitive states (although contemporary blog writings may lend weight to the account). Lifelog evidence may be legitimately challenged (just because one's GPS unit was at point x at time t, that does not mean the lifelogger was the person carrying it). One may of course be guilty all along, and the lifelog may provide evidence against one, a possibility against which Allen warns (2008). The existence of a lifelog, if known, would render it liable to seizure and examination by the authorities.

Accountability of others

The technology is also 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 (cf. Bailey & Kerr 2007, Brin 1998, Mann & Niedzviecki 2001, O'Hara & Shadbolt 2008, 181-183).

There are obviously pros and cons to sousveillance, and many have drawn attention to the privacy concerns therein, but an indiscriminate account of one's own life is *ipso facto* an indiscriminate account of one's dealings with authority, and can therefore be used to hold them to account. Typically one would expect to have to bring information from the lifelog into juxtaposition with other information to provide significant accountability, and the possibility of this will increase if the lifelog is furnished with a rich set of metadata.

The empowerment envisaged here depends on one's identification with a community. It is the community that will benefit from, e.g. the exposure of a political representative as neglectful of constituents' concerns, and it is the community as a whole that will need to provide evidence of that neglect, and indeed to publish and read it. Assuming that the lifelogger is interested in his or her community welfare, then by the provision and collection of lifelog information he or she is empowered to make a contribution.

Discussion

The previous section remains speculative, both in respect to whether lifelogging will seriously reduce privacy, and whether it will be empowering. There are grounds, we suggest, for thinking that the good will outweigh the harm, and indeed that the harm will be minimal. Ultimately, this is an empirical matter, as all authors on the topic freely admit, and our speculations may well need to be revised once hard data comes in and tricky cases hit the courts. At a minimum, however, we argue that there is no reason to rush into regulation.

As lifelogging tools are currently a matter for research rather than commercial application, there is a window for reflection about privacy and regulation. There have been some unrealistic suggestions, such as programming imperfections into the system, deliberate random error that would prevent the lifelog from being veridical, and therefore from invading privacy (Dodge & Kitchin 2007). There has also been unfounded optimism, as with Cheng and colleagues, who are sanguine about the use of lifelog data by the judicial system (Cheng et al 2004). In contrast, Allen warns that lifelog data would be fair game for most governments, and recommends that people

should never be required to keep a lifelog, should own their own lifelogs and should be allowed to delete content at will (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. One should not be 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 are necessarily personal, though they may be public domain, but why think that information gathered should be kept private? The publication of the information might just be the point.

Furthermore, although Mann, Elahi and Ringley relish or relished the unblinking gaze of the technology, and Bell is omnivorous in his information consumption, the universality of their approaches is (in the cases of Mann and Bell at least) facilitated by large-scale resources behind them, and in all four cases licensed by a personal ideology of information. Real life, whether online or off, is rarely so ideologically driven or well-resourced, and it is unlikely that many lifeloggers would be universalist in their assumptions. A more likely scenario is that the information stored would be rich in some respects and poor in others. The logger might invest in a new gadget that provides information of a certain sort in detail, before the novelty wears off and the information dries to a trickle. Loggers might periodically forget to take GPS devices with them on their perambulations. New interests or a change of circumstance might result in more or less information being stored. The picture is likely to be patchy; undiscriminating information gathering does not entail an unblinking gaze.

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.

There may still be worries about the permanence of the record. Transience is one of memory's outstanding aspects and forgetting is part of good mental housekeeping (Schacter 2001). Nevertheless, the picture is not as clear as might be thought. Lifelogging can be used to measure which pieces of information are recalled directly or used indirectly in associative recall, which could help structure search and recall so that information that was never or rarely used was less prominent in results (rather as PageRank imposes an ordering of importance on webpages). Paradoxically, one could use lifelogging techniques to decide which information to delete (although deletion goes against the grain of the lifelogging ethos).

Conclusion

Commentary on lifelogging has tended either to geeky techno-optimism, or warnings of potential dangers. The optimism is probably overdone, as optimism tends to be. Certainly the dangers exist, but the discussion so far is framed on possibly false assumptions that lifelogs will (a) consist of personal information, (b) be universal in

scope, (c) include information that has traditionally been held private by owners, and (d) become a mainstream activity, possibly via social pressure. The falsity of any one of those assumptions would undermine the arguments against lifelogging, and it is quite conceivable that all four of them are false.

That is not to say that privacy questions will not arise with lifelogs, as they clearly do with other digital activities such as social networking. Reasonable expectations of privacy are currently in flux – the Web is evolving faster than legal and ethical systems can respond – but lifelogs do not pose *special* or *unique* issues, despite the undiscriminating nature of information gathering.

We have also argued that lifelogging can be empowering for the logger, allowing him or her information which can be used in the construction of a personal online identity, or identities, which is not under the control of authorities. Furthermore, as accountability is increasingly important in society, lifelogging can help both in accounting for the lifelogger's personal behaviour, and in holding others to account.

A final point: lifelogging sounds somewhat recondite, 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

SIOC: Semantically-Interlinked Online Communities

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

UKCRC: United Kingdom Computing Research Committee

URI: Universal Resource Identifier

Acknowledgements

Our thanks to the audience at the workshop on identity in the information society in Arona, Italy in 2008, and for rigorous comments from two anonymous referees.

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 (1988). *Uneasy Access: Privacy for Women in a Free Society*, Totowa NJ: Rowman & Littlefield.

Anita L. Allen (2003). 'Privacy isn't everything: accountability as a personal and social good', *Alabama Law Review*, 54.

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

Jane Bailey & Ian Kerr (2007). 'Seizing control? The experience capture experiments of Ringley & Mann', *Ethics and Information Technology*, 9, 129-139.

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

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.

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.

Luciano Floridi (2005). 'The ontological interpretation of informational privacy', *Ethics and Information Technology*, 7, 185-200.

Jennifer Golbeck (2006). 'Trust on the World Wide Web: a survey', *Foundations and Trends in Web Science*, 1(2), 1-72.

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.[18]

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

Dan Jones (2008). 'How to protect your good name against cyberspite', *New Scientist*, 22nd May, 2008.

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

Paul B. de Laat (2008). 'Online diaries: reflections on trust, privacy, and exhibitionism', *Ethics and Information Technology*, 10, 57-69.

Stew Magnuson (2007). 'Army wants to make "every soldier a sensor"', *National Defense*, May 2007, http://www.nationaldefensemagazine.org/archive/2007/May/Pages/ArmyWantSensor2650.aspx?PF=1&PF=1&PF=1.

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

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

Mor Naaman, Yee Jiun Song, Andreas Paepcke & Hector Garcia-Molina (2004). 'Automatic organization for digital photographs with geographic coordinates', in Hsinchun Chen, Howard D. Wactlar, Ching-chih Chen, Ee-Peng Lim & Michael G. Christel (eds.) *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, Association for Computing Machinery, 53-62.

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.

Madeleine Sorapure (2003). 'Screening moments, scrolling lives: diary writing on the Web', *Biography*, 26(1), 1-23.

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.

Daniel J. Weitzner (2007). 'Whose name is it anyway? Decentralized identity systems on the Web', *IEEE Internet Computing*, May/June 2007, http://dig.csail.mit.edu/2007/06/ieee-ic-decentralized-identity-weitzner.html.

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.