Towards Sharing Life-Log Information with Society

Reza Rawassizadeh
Department of Distributed and Multimedia Systems, University of Vienna, Vienna, Austria
rrawassizadeh@acm.org

Abstract

We are living in an era of social medias such as online communities and social networking sites. Exposing or sharing personal information to these communities has risks as well as benefits and there is always a challenge between the risks versus the benefits of using these technologies. Life-logs are pervasive tools or systems which sense and capture contextual information from the user’s environment in a continuous manner. A life-log produces a dataset which consists of a continuous streams of sensor data. Sharing this information has a wide range of advantages for both user and society. On the other side, in terms of individual privacy, life-log information is very sensitive. In this paper, first we list risks and benefits of sharing life-log information. Although social medias enable users to define share limitation for their information, but because of the life-log data structure, current sharing models are not capable of handling life-log information. Therefore, based on the identified risks we propose a data model for sharing life-log information. This data model has been designed for life-logs and it tries to reduce the associated risks. Furthermore, ethics for providing and using life-log tools will be discussed. These ethics focus on reducing risks as much as possible while sharing life-log information.

Keywords: Life-Log, social computing, information sharing, personal information, privacy

1 Introduction

For the majority of us the internet is a necessity in our daily life. Lots of our personal information is available online and is easily accessible to other users via the social oriented technologies such as social networking sites (SNSs) or online communities. Users benefit from using features of these society-aware technologies by sharing their personal information with other users. For instance a user can share content, find new friends based on the shared content (common interest) and stay in contact with old friends, etc.

These technologies allow users to enrich content via different mechanisms such as ranking, tagging, commenting, etc. In simple terms, content will get enriched by sharing. Most famous content sharing sites such as Youtube\(^1\) and Flickr\(^2\) offer social network functionality to use crowd annotation and enriching their content.

Berslin and Decker [Berslin and Decker(2007)] predict that social networking will go beyond ego surfing in the future. They introduced a social networking stack to let users share information beyond the SNSs domain, e.g. desktop environment. Based on their prediction we can conclude that all user’s digital objects and documents can be shared with society, with enough consideration to the access control limitation. On the other hand, exposing personal information outside private space increases the risk of misuse by criminals, governmental institutions, businesses institutions, and any other third parties. Social computing environments lack governance by their nature [Parameswaran(2007)] and potentially these environments are very prone to misuse. There is always a compromise between sharing information and considering the privacy of the users. There is also less research done on privacy in social communities [Nov and Wattal(2009)].

In 1945 Vannevar Bush [Bush(1945)], described an imaginary device (memex) in his famous article called As we may Think. This device can record and link books read by the subject, movies watched by the subject, etc. Now in the era of digital technology Bush’s idea has been interpreted as life-log. Life-log tools and applications mostly focus on augmenting the memory of the user. They capture the user’s actions and contextual information. The life-log produces a dataset from this information. This

\(^1\)http://youtube.com
\(^2\)http://flickr.com
dataset is semantically very rich and provides lots of
information about the user’s context and situation.
User and society can both benefit in many
ways by sharing life-log datasets with society
such as story telling, historical studies, crowd
health monitoring and group behavioral studies
[Eagle and Pentland(2009)]. We believe these po-
tential advantages will lead to more business and
scientific attention in the near future. There are
some art projects [Dodge and Kitchin(2007)] which
have exposed life-log information to the public,
but to our knowledge there are fewer scientific
efforts being made in this area. Reality mining
[Eagle and Pentland(2006)] is one of the pioneer
projects which has this feature, it focuses on studying
social behaviours of individuals in a group.
Sensitivity of a shared information depends on
the information object. We call any readable or
understandable data block in the life-log dataset an
information object. Some information object are
less-sensitive such as a book’s rank or a movie’s
rank, some are more sensitive, such as email address,
and some are highly sensitive, such as life-log data
which contains information about the private life of
the user.
In this paper we try to list potential risks and
benefits of sharing life-log’s information. Based on
the identified risks, we propose a data model that
enables the user to share his life-log’s information
with society. In order to reduce associated risks,
this model will provide appropriate access limitation
to user’s life-log information. It defines privacy and
access scope for the smallest information object.
Smallest information object here means atomic
information block such as an email address, location
at the specific timestamp, etc. A fine grained access
limitation model is not enough for reducing risks,
therefore we propose some additional ethics for
providing and using life-log services. Proposed ethics
could increase trust between the user as the service
consumer and the service provider. Basic level of
trust will be required to motivate individuals using
these services.
The remainder of this paper organized as follows:
Next section identifies and describes risks and ben-
efits associated with sharing the life-log data with
society. Afterwards based on the identified risks,
a data model will be introduced to define access
limitation on any information object in the life-log
dataset. Then additional ethics which are necessary
for dealing with life-log services will be described.

At the end we conclude this paper.

2 Risks and Benefits
In order to be able to analyze the cost of sharing the
life-log information, we classify the risks and the ben-
efits separately. Here we identify risks and benefits
mostly based on theory. This means it is possible
that in the future more risks or benefits will be intro-
duced in this domain, because few experiments have
been done with the real operational system.

2.1 Benefits
Sharing life-log information could benefit individuals
in educational, business, health cares and social rela-
tions. Some of the identified potential advantages of
exposing life-log data to society have been listed as
follows.

- **Software Personalization**: Software applications
can be personalized based on interests of the
group (social interest). Software personaliza-
tion could facilitate search and information re-
trieval. Studying and mining group preferences
can be done based on the past activities (life-
log dataset) of the group. For instance, when
a user searches for “ant” in a software develop-
ment environment, it means “Apache ANT”
http://ant.apache.org and not the insect ant.
In simple terms, applications can be personalized based on the groups
characteristic and group characteristic is identi-
fied by studying life-log datasets of group mem-
bers.

- **Learning Social Patterns and Behavior**: Life-log
data can be used to learn the social behavior of a
group. For instance in the case of earth quakes,
location changes of the individuals can be be
logged by the location sensor of their mobile
phones [Eagle(2008)], to assist scientists study-
ing how they can reduce the damage.

- **Matchmaking**: The first requirement to start
a new social relationship is to have a common
interest. Common interests between users can
be extracted from their life-log dataset. Reality
mining [Eagle and Pentland(2005)] perform

3http://ant.apache.org
matchmaking based on the user profiles and behavioral data of the user. There are more potential commonalities, which can be used to suggest a match, such as the user’s habits (e.g. eating habits of the subject can be extracted from locations of visited restaurants).

- **Recommendation Systems**: A recommendation system can study and extract knowledge from the life-log dataset of the users. For instance, place ‘A’ has been visited more than the place ‘B’. A recommendation system can suggest that the place ‘A’ is worth visiting more than the place ‘B’.

- **Health and Medical Studies**: Biological sensors can be used on a subject’s body to sense and record biological data such as body temperature, skin conductance, body heat, etc. This data can be recorded and used as the life-log information. The health status of society can be monitored via provisioning and monitoring life-log information of the individuals. Sharing personal health information with society can assist prevention or determination of an epidemic. O’Hara et al. (2009) identified two important potential hazards of life-logs: pernicious memory and pernicious surveillance. In order to be able to provide a data model and identify ethics that reduce the potential hazards, we describe them here in a different classification and we try to identify more risks. As with benefits, we cannot argue that we list all potential risks. More risks may be identified, when a life-log tool is used in a real operational environment and for business purposes.

- **Historical Studies**: Memory augmentation is the main goal of life-logs. When life-log data is shared, it can assist individuals in sharing their life events. History composed from the life events of individuals. Maintaining a life-log dataset in the long term is a valuable source which might be used for historical studies.

- **Sousveillance**: This terminology is different than surveillance. Surveillance means recording the individuals activities by technologies on behalf of an organization. Surveillance systems are not hazardous by default, they support safety, welfare, health, efficiency, speed and co-ordination. Martin Dodge and Rob Kitchin (2003) described that a life-log could be interpreted as a form of personal sousveillance which is bidirectional, not like surveillance unidirectional (government monitor individuals). Sousveillance can democratize the process of surveillance via supporting the monitoring of the authorities (surveillant).

2.2 Risks

As has been described in the introduction, sharing personal information has risks and benefits; Losing privacy is the most important risk of sharing personal information. Unfortunately life-logs have a controversial history, such as a DARPA’s lifelog project which was canceled in 2004 because of criticism of the system’s privacy implications. The Life-log dataset contains sensitive personal information such as the user’s location, which requires more privacy concerns, because sharing personal information with this level of detail can be hazardous. Anita L. Allen (2007) identified two important potential hazards of life-logs: pernicious memory and pernicious surveillance. In order to be able to provide a data model and identify ethics that reduce the potential hazards, we describe them here in a different classification and we try to identify more risks. As with benefits, we cannot argue that we list all potential risks. More risks may be identified, when a life-log tool is used in a real operational environment and for business purposes.

- **Surveillance**: Sharing life-log information with society might be interpreted as a form of surveillance. The surveillance, which limits our behaviors without our desire, is not acceptable. Surveillance has some potential disadvantages such as increasing the number of suspects who are not guilty. Security agencies, governmental organizations, business organizations, criminals and other types of organizations or industries which are benefit from social control can misuse surveillance data. For instance a recruitment company can check a candidate’s psychological and physical health and his life style during the recruitment process, in order to find out that if the candidate is suitable for the offered job or not. If the candidate performs extreme sport, the chance of having time off for medical purposes will be high. Existing laws and policies do not provide an appropriate limitation on the unwanted use of personal information. In the U.S. CAELA (Communications Assistance for Law Enforcement) enforces communication technologies to allow government access and surveillance. Also, in March 2006, the European Union adopted a Directive, which mandates that the content of electronic communications services will not be deleted (remain for
not less than six months and not more than two years). They can be used for marketing and provisioning purposes [dir(March 2006)]. Observing users in a community without their awareness is the worst form of surveillance. This form is long-standing. First time Jeremy Bentham (1838) defined it as Panopticon, then Michel Foucault [Foucault(1977)] compared Bentham’s ideas with modern society.

- **Memory Hazards:** It has been approved that life-log tools can assist human memory [Sellen et al. (2007)]. The life-log can record all life events, disregarding the content. It means a life-log can prevent individuals from forgetting their errors [Dodge and Kitchen(2007)]. Humans naturally try to distance themselves from their errors and misfortunes because, psychologically, humans need to forget misfortunes. Life-logs could be harmful in this case and they can also cause pathological rumination for unipolar and bipolar depression [Michael E. Addis(1999), Allen(2007)]. Exposure of personal mistakes to society might be more harmful than keeping them private. For instance children by their nature are weaker at bearing misfortunes, they can watch parents life-log and remember a misfortune which had happened in their family. Another problem can appear by sharing the mental problems of individuals, which could have a negative impact on crowd mentality in a team or a group of individuals sharing a common interest. To handle memory hazards of life-logs M. Dodge and R. Kitchen [Dodge and Kitchen(2007)] suggested that the life-log should forget like real memory, based on the Scharter’s six forms of forgetting [Schacter(2001)]. We believe fading life log memory will distance it from its main goal as an ultimate memory aid, because we will lose the reliability and persistence of digital memory.

- **Long term Availability of Personal Information:** The ideology and personality of an individuals can change over time. Also an individual’s lifestyle may change over time. Sharing life-log information with society can be a permanent record of our mistakes. For instance, imagine a group of teenagers gathered at a party and posed for the camera in an embarrassing way. These pictures go online and every body can see them. At that time there is no problem. 30 years later one of those teenagers is now going to participate in an important political campaign, these pictures could harm his career. These problems are not only for the life-log information, other online traces of the individuals can cause him to suffer also. Taking another example, consider a writer who has changed his mind over time and disagrees with his past opinions. In the era of the internet the chance of removing his old ideas from public view is very low.

- **Stealing Life-log Information:** Sharing life-log information increases the chance of stealing or losing it. This risk can be very harmful for the person (victim). Lior J. Strahilevitz [Strahilevitz(2004)] claimed most private information consists of sensitive personal matters such as sexual encounters and bodily functions, sensitive medical information, knowledge of owner’s fundamental weaknesses, etc. The life-log can sense and record this information, therefore life-log’s dataset is a very sensitive object from a privacy point of view.

The described risks indicate that life-log could be harmful as well as being useful. Except memory hazards all other risks are related to the sphere of privacy. Therefore, any associated data models for life-log tools must carefully manage privacy related issues.
3 Model Definition

Securing privacy is the greatest challenge in using social networking systems. Trust is the basis of privacy, and is established based on different factors e.g. social relation. This means first we need to identify individuals’ roles in this domain. Furthermore life-log’s data structure must deal with the social networking features of these tools. Based on identified roles and life-log data structure we propose a data model for sharing life-log information. This model is able to define access limitation on the smallest information object. We focus on defining the conceptual data model to have a technology independent solution.

3.1 Roles in a Social Community

We identified three roles in the context of a social community. The end user is the person who benefits from the system by exposing his personal information to the community. The community owners are people who own the social community and benefit from having users and probably use their information. Increasing the number of users, raises the community value. The third parties are groups or individuals who benefit from communities by using a community environment to advertise, perform research on the community, study user behavior, monitor community member activities, etc.; they are able to access user’s information in the community with or without the end users’ awareness; they could vary from marketing businesses, security agencies to hackers. In order to use the user’s information, third parties must make an agreement with community owners. Community owners usually inform users about this in their “Terms and Condition” section of the community’s membership agreement criteria. Users share their information with firms and other community members [Nov and Wattal(2009)]; they share information in the community based on trust. From the user point of view third parties have the least credibility in the community; they observe users’ activities in the community and benefit from the community. They can also be the business focus of the community owners. In simple terms, community owners benefit from providing information of the community to the third parties.

3.2 Life-log Data Structure

Life-log tools are used to capture user’s daily life events. The information can be interpreted as a subset of personal information. A life-log contains a dataset of a subject’s life events. Each life event of the subject is a record in this dataset. In our proposed data model, each life event (life-log record) has an access scope. These tools use sensors to read and capture contextual information. In simple terms, they acquire data from reading sensors, information flows to the life-log from multiple sources (sensors) in a continuous manner. There is no guarantee for the sequence of received data and the sequence varies over time. Also, life-log dataset size is unbound. All these properties indicates that life-log dataset is a kind of data stream [Babcock et al.(2002)Babcock, Babu, Datar, Motwani, and Widom].

We are living in a spatio-temporal world. Meanings, all of our life events except dreams happen in a specific location and at a specific date-time. Based on the current available technologies it is not always possible to sense the location, because location sensors such as GPS are not functional in every environment. For instance GPS can not work indoor. There are other approaches such as A-GPS (Assisted GPS) to solve this problem, but they can not always sense the location and they are imprecise. On the other hand most operating systems have date-time which is accessible since the target device has not been turned off. This means most devices with computing capabilities can provide timestamp. Therefore we conclude that date-time is a necessary field for any life-log record and all life-log information objects will be stored with the timestamp. Life-log data types vary based on the sensor output. Data can be a text or a binary object such as a movie or a picture. We assume all records of a life-log dataset have a date-time without considering the data-type of the record. Life events can be recorded in a continuous-time manner or discrete-time manner. Audio files and video files are examples of continuous data. In simple terms, if there is a start time and an end time for an event then it is a continuous-time event and when there is a unique time, then it is a discrete-time event. Our proposed data model has no dependency on the type of life event (discrete or continuous). Continuous-time data can also be called time series data too [Gibbs et al.(1994)Gibbs, Breiteneder, and Tsichritzis].

3.3 Data Model

An access scope will be defined for each life event. As previously described, each life event is a record in
the life-log dataset. Atomicity is the main property of a life event. The dataset composed from a set of infinite life event data objects, which come from heterogeneous sensors. The dataset size is infinite because the user’s life is ongoing and during his life the dataset size increases. Also there is a wide variety of sensors that can be used by the life-log and they provide different data types with different structures. As it has been noted before this dataset is a data stream. The proposed model should not have any dependency on sensor’s data-type or data structures. In other words, there should not be any boundaries between the access scope definition and the sensor structure.

We provide a conceptual definition of the model, which enables users to share their life-log information. It is not dependent on technology and a life-log dataset can be in any data format such as XML, JSON[4] or RDBMS[5].

Information objects in social communities are time independent, this means that they do not support time based access limitation, but life-log data is date-time (timestamp) dependent. Therefore, we use timestamp data as a mandatory field for each life event record. It is notable that in the life-log domain nobody other than the owner of the dataset is able to manipulate the data and the only access right is “read”, nor any other data manipulation rights.

Each life event is a data entity or record in the dataset. The Life-log dataset of a person $P$ represented via $L(P)$ and $L(P) = \{E_1, E_2, ...E_n\}_{n \rightarrow +\infty}$.

The life-log data entity $E$ is a 3-tuple $(D, T, S)$, where $T$ is the timestamp, which can be continuous or discrete; if continuous, it will be the start timestamp and the end timestamp. $D$ is the information object and it can be binary data e.g. image, audio or textual data e.g. GPS location, micro blog content, etc. $S$ is the access scope. We defined three access scopes, *Private*, *Public* and *Friend*. Gross et al. also Gross et al. (2005) Gross, Acquisti, and Heinz identified these three scopes and named them, private, semi-public and open-ended. Private means that the data entity is not shared and nobody other than the owner can access this information object. Public means that everybody in this social domain can access this information object and no access limitation has been defined for it. *Friend* defines users who can access this information object. The user can define which user(s) or group of users from the social domain can access this object. $S = \{P_1, P_2, ...P_n\}$, $S$ is a finite set of users or group of users who can access user’s information objects. A social relation description models such as FOAF[6] can be used to extract the list of users or the users who can see this data entity or it can be specified manually by the user. How to extract the user list is not in the scope of this paper. Developers or end-users can define it based on their interpretations of the system.

Please note that only one access right granted to users which is the “read” right. Unlike other access control systems we do not have other permissions such as delete, update, etc. It is the authorization process and not the authentication process, because only “read” right on a information object might be granted to the user. It is tedious for the user to specify the scope for each data object, therefore in the implementation model he can grant access to a user or a group of users for a group of objects, based on time. For instance all videos from the beginning of March up to the end of April can be marked to be shared in a family scope which is a group of users.

In order to hold the granularity of the model, the timestamp can not be the unique identifier for defining the share scope, it should be used with the data object. For instance consider Figure 1, which shows that three sensors have been used. The user intends to share his video between the time $t_3$ and $t_4$, and not his heart beat rate. Thus share scope must be defined based on the timestamp and the data object and not the timestamp alone. This assumption leads us to make each data entity record a 3-tuple.

Elements arrival order to the $(L(P))$ dataset are based on the timestamp, but there is no guarantee. For instance, in order to reduce the processing cost, a buffering mechanism can be used for reading a sensor. It will gather sensors’ data and when the buffer size reaches a specified size then the data will be sent to the dataset.

Figure 1. Video sensor is sampling in the continuous-time manner and other sensors are sampling in the discrete-time manner.

This model is a conceptual approach to enable the user to share his life-log’s information with society. It is designed to be independent of implementation technologies and it is flexible enough to allow

---


[5] Relational Database Management System

developers manipulate data entities, but we think these three fields are necessary components for each data entity. Implementation of the model and how to restrict access or enable access is not within the scope of this paper. During implementation more questions and challenges might be raised such as in order to share bulk data e.g., video how to store them on the server? Is a link to local storage enough or do we need to upload whole data to the server? In addition to access control, other security aspects of the system must be taken into account during the implementation of a social aware life-log. Communication and data transfer are very sensitive from a security point of view and sharing information with society requires communication. This emerges that during the communication, data should be transfer encrypted. Securing the connection can be done via the transport layer security (TLS). However additional to the TLS, using digital signature has been suggested as an additional way to secure the data transportation.

4 Ethics for Sharing Life-Log Information

In the last section we discussed how a user can share his life-log information, which is very sensitive, with society and we have listed associated risks and benefits. Another important thing which remains undiscussed is the related ethics for providing and using these kinds of services because just having an appropriate security mechanism and data model is not enough. Service providers and service consumers (users) both need to apply ethical considerations while dealing with these features. Anita L. Allen [Allen(2007)] discussed some important ethical issues concerning privacy in the life-log. Based on her efforts and our interpretation, we will try to identify ethics and list them here explicitly. It is possible that some points are missing from our list and during real system usage in a business environment more ethical considerations will be discovered.

User’s rights and responsibilities could be listed as the following:

- The user is not mandated to keep his life-log data; he is able to delete, edit or add content to his life-log dataset. In other words, the user is able to change his data anytime he likes without any requirement to provide a reason to the service provider.
- The user should not share information about other users without their permission or legal guardians permission. For instance appearance of a third person in a photo and annotating him in that photo by the user is prohibited without his consent.
- In the case of misusing a third party’s information which has been extracted from the user’s dataset, the user is also legally responsible.
- Any illegal misuse of information by the user is the responsibility of the user and not the service provider.
- The user must be aware of any data manipulation policy. For instance the service provider must clarify whether the deleted data will remain in the server backups or not if the user deletes his data.
- The user should be able to enable or disable a sensor. It means that the service provider must provide enough features to let users configure sensors based on their preferences. For instance, the user decides to log his location and not his video on day A. Next day, on day B, he decides to log both.

Based on the identified risks we recommend that service providers must clarify the following with service consumers when an agreement between two parties is made:
Life-log data is the property of the person who creates it. It means the user is one of the owners of this data and service providers which are bound by ethical issues should give users the right to manipulate and manage their data. Additionally ownership of each user’s data component must be clarified. Either the user is the sole owner of the data or the service provider and user both are owners. The data component can be an information block such as the user’s email address or it can be sensitive information about the user such as the user’s biological information.

The user’s data cannot be manipulated by the service provider without the user’s consent. Except in the case where the user intends to commit fraud, misconduct the service provider or any other usages which break the agreement between the user and the service provider.

The service provider should clarify its rights in using the user’s data, and the user can accept or reject them during the service registration process. We suggest that the service provider should not be allow to copy or transfer the user’s private data to a third party without the user’s permission.

The service provider must provide counter technologies such as encrypting user’s information, providing authorization and authentication, etc. to block unauthorized surveillance of the user.

The service provider must define its’ responsibilities in the case of data corruption, eavesdropping or any form of data loss and data misuse.

The service provider should inform users about third parties who intend to use their data such as scientific institutions, security agencies, etc. User awareness is not required in every case, e.g. observing visitors in a street, since it does not focus on any individual’s privacy, is not a problem.

The service provider must clarify that if after the user’s death his information could be inherited or not. If yes by whom can this information be inherited? If not, what is the service provider’s policy about the user data?

In social networking systems some issues are not clear, such as data ownership after the user’s death. To our knowledge Legacy Locker [10(2009)] is the first effort to manage the problem of online data after death of the owner and some social networking systems have a solution for the deceased user, e.g. Facebook follows a family’s wishes to take down a deceased user’s profile or keep it in a “memorial state” [11(2009)]. How harmful is the stolen life-log dataset for the user? What is the impact of a life-log on a society? These are some open questions which should be answered when a real business oriented system is available and operational.

5 Conclusion

Here our focus is on analyzing different aspects of sharing the life-log information. First we identified risk and benefits to indicate how important it is to have a specific privacy aware model for life-log tools and sharing life-log data. Then we proposed a data model which defines access scope for the subject’s data (user’s life-log data). This model is flexible enough to define access scope on any information object in the life-log dataset disregarding the data type and the data format of the data object. Having a fine-grained access model is not enough because life-log’s data is very sensitive from a privacy point of view; therefore we introduced some ethics which can assist user’s safety while sharing his life-log information. Social network systems are rather new technologies and probably far from maturity. Life-logs are also more available in the scientific domains than real operational systems. Our goal for writing this paper is to open a discussion on sharing this information with society. It is a wide open scientific area and this effort is just a small step.

Acknowledgment: We acknowledge Prof. Wolfgang Klas from the department of Distributed and Multimedia Systems at the University of Vienna and Dr. Shahin Gheitanchi from the Communications Research group at the University of Sussex for their helpful hints and feedback on this research.

References


[O’Hara et al.(2009)] O’Hara, K., Tuffield, M., and Shadbolt, N.,


