Social Science Research Approach for Lifelog Utilization

Katsumi Takahashi†

Abstract

This article describes privacy issues related to the use of lifelogs. After mentioning legal systems for privacy, it introduces privacy preserving data utilization technologies. It also describes research trends for major techniques such as data anonymization and secure function evaluation. Some important aspects of lifelog utilization are cooperation among service design, legal system preparation, and privacy preserving data utilization technologies.

1. Introduction

Can lifelogs be used to improve services? My colleagues and I believe that lifelogs are useful in several daily situations. For example, you may have experienced getting a good book recommendation from an online bookstore and some people would be interested in pedometer records on their mobile phones to monitor their health. These services can be provided automatically by collecting and analyzing records of actions, which are called lifelogs. The mechanism is based on the concept that collecting more action records enables more useful information to be found in them. This mechanism requires precise action records from other people. However, if we collect more action information, privacy issues arise. We should avoid collecting action information that is identified with specific people because that would represent a breach of their privacy. We are the first generation whose lifelogs are being collected and utilized. Nobody knows the proper way to use lifelogs yet.

Who can answer the question of how to use lifelogs properly? First, the end users who provide the lifelogs must give consent. Second, service or business model issues that provide merits for both users and service providers should be discussed. Third, continuing lifelog utilization requires legal systems to show the standard or adjust interests. Such discussions have led to the development of some technologies, which are collectively called privacy preserving data utilization (PPDU) technologies. Establishing lifelog-based services requires a cooperative approach among useful services, legal systems, and PPDU technologies (**Fig. 1**).

NTT Information Sharing Platform Laboratories is interested in personal data utilization technologies and is discussing the issue of technologies and social systems. This article describes trends of legal systems and PPDU technologies.

2. Legal systems for lifelogs

In Japan, there are currently no laws or guidelines directly related to the use of lifelogs. We must determine the basic idea from the traditional privacy protection scheme. OECD (Organization for Economic Cooperation and Development) published "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data" in 1980. It recommends the way to handle personal data including, collection limitation and purpose specification principles. These recommendations have become the basis for the personal data processing of each country, including Japan.

OECD's recommendations are shown in **Fig. 2**. The "Act on the Protection of Personal Information" (Act on PPI) was worked out in 2003. In this law, personal information is defined as shown in **Fig. 3**. "Proper

† NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Tokyo

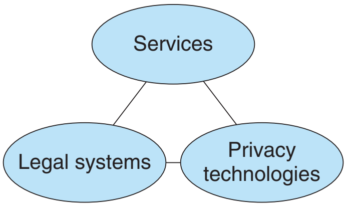


Fig. 1. Triangle for lifelog utilization.

1. Collection Limitation Principle
2. Data Quality Principle
3. Purpose Specification Principle
4. Use Limitation Principle
5. Security Safeguards Principle
6. Openness Principle
7. Individual Participation Principle
8. Accountability Principle

Fig. 2. OECD's privacy principles.

Article 1 Definitions

The term "personal information" as used in this Act shall mean information about a living individual which can identify the specific individual by name, date of birth or other description contained in such information (including such information as will allow easy reference to other information and will thereby enable the identification of the specific individual).

Fig. 3. Definition of personal information.

acquisition" (Article 17), "Notice of the Purpose of Utilization" (Article 18), and "Restriction of Provision to a Third Party" (Article 23) are defined.

This law has clarified what should be observed when handling personal information. Lifelog handling is often regarded as personal information handling, but there is discussion that the Act on PPI does not help lifelog utilization although "consideration of the usefulness of personal information" is described in the purpose chapter (Article 1). The discussion includes "Does the Act on PPI catch up with new market needs or social needs?" and "what kind of lifelog corresponds to personal information?"

The "Information Grand Voyage" project of the Ministry of Economy, Trade and Industry (METI) from 2007 to 2009 investigated personal data utilization. The main technological idea for safety is the anonymization of identifiable personal information. A study for best practices has been conducted and the software platform for anonymization has been developed [1]. A Ministry of Internal Affairs and Communications (MIC) study group that has been discussing issues related to information and communications technology (ICT) services from the user's standpoint since 2009 has reviewed lifelog services and made some proposals. The proposal documents request the following: "A lifelog is a broad notion that includes web access histories, payment histories, and location histories. Although information for targeted advertisements does not require personal identifiable information, if it can be identified by looking up other data, or inferred from data collected over a long term, it should be regarded as personal information." [2].

Considering social trends and needs, the approaches of both METI and MIC encourage the drawing up of independent guidelines rather than the designing of controlling institutions. We believe that any approach requires operations that use privacy protection technologies.

3. Privacy preserving data utilization

In developing services that use lifelogs, one must take care to ensure the privacy of the persons providing the lifelogs. In the same way that the balance of a bank account is shared only by the account holder and the bank, if a lifelog is shared only by the system and used for the person, the usual security control works well. But lifelogs may be shared by several systems, so privacy is important. A framework of lifelog processing is shown in **Fig. 4**. First, the raw lifelog provided by a certain person should be collected ((a) in Fig. 4). Second, there are issues concerning the processing of collected lifelogs, such as storing and analyzing them (b). Third, after analysis, the analyzed statistical data might require privacy (c). PPDU is a technique that preserves privacy throughout these steps.

PPDU covers three methods [3] (**Fig. 5**). In this article, we describe the classic data anonymization

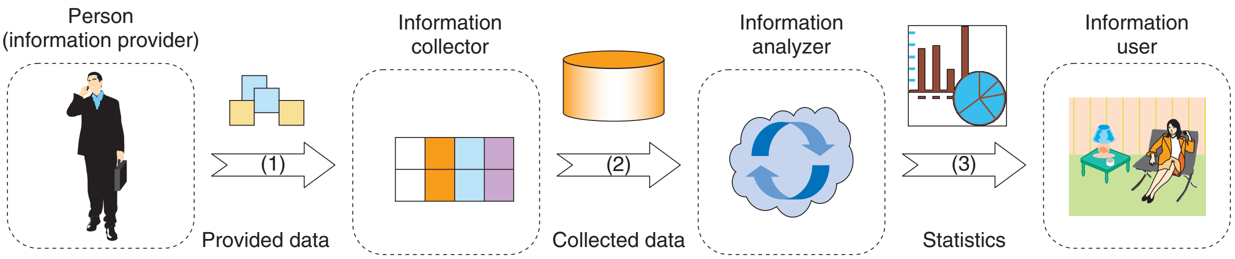


Fig. 4. Lifelog handling process.

Anonymization: Generalizing data so that it cannot be identified
(example) Ginza, Chuo, Tokyo; male; 33 → Chuo, Tokyo; male; 30s

Perturbation: Adding noise to generate perturbated pseudo data
(example) Ginza, Chuo, Tokyo; male; 33 → Ginza, Chuo, Osaka; male; 35

Secure function evaluation: Analyzes encrypted data without revealing the raw data
(example) Ginza, Chuo, Tokyo; male; 33 → %52TE4AS &HR*YS S!A@3S

Fig. 5. Three types of PPDU.

and the development of an improved secure function evaluation.

4. Data anonymization

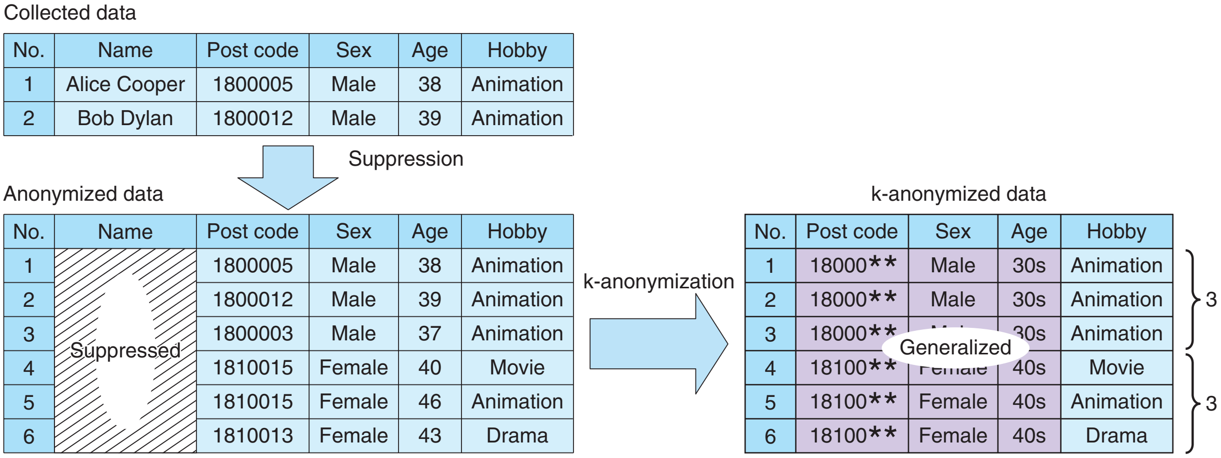
If you wish to share a large amount of personal data (such as lifelogs), anonymization, it is a good idea to make data unidentifiable. The basic method for doing this is anonymization, which is a method of modifying the data to make it unidentifiable while leaving as much useful information as possible. There is an empirical method of suppressing identifiers such as names and addresses. However, even if names and addresses are deleted from the data, some data could be identified if it contains unique hobbies or actions. Such data is not safe in terms of privacy. On the other hand, if you generalize the data too much in consideration of privacy, the data will be less useful. What is required is an anonymizing method that controls data for both safety and utility. A data anonymization technology and a method for efficiently dealing with this trade-off have been studied.

One major measure is k-anonymity. This represents the data status that there are at least k records that share the same combination of attributes (which might be used for identification, such as post code, sex, and age, as shown in **Fig. 6**). Anonymization software that produces k-anonymized data efficiently has been developed in the Information Grand Voyage project.

5. Secure function evaluation

Another method for processing more sensitive data is called secure function evaluation (SFE). This technique analyzes the data under a prior agreement without revealing the raw data to anyone involved in the whole process. Ordinary cryptographic techniques must decrypt encrypted data in order to analyze it. But NTT's SFE, which we have been developing, divides the original data into several fragments, processes the data fragments, and restores only the calculated result to a visible format.

A simple example of SFE is shown in **Fig. 7**. It shows the process of calculating the average score of three children without revealing their raw scores to anybody else. First, each student splits his or her score into two parts and sends each part to a different server. The two servers calculate the average of all the fragments they receive and these two average scores can output the average score of the three children. SFE can provide not only this elementary sum function, but also four arithmetic operations and logical operations including accordance and numerical comparison. By combining these operations, SFE can perform complex processes such as providing statistics, data matching, and data mining. SFE requires two or more individual computing agents, as



Collected data

No.	Name	Post code	Sex	Age	Hobby
1	Alice Cooper	1800005	Male	38	Animation
2	Bob Dylan	1800012	Male	39	Animation

Suppression

Anonymized data

No.	Name	Post code	Sex	Age	Hobby
1	Suppressed	1800005	Male	38	Animation
2		1800012	Male	39	Animation
3		1800003	Male	37	Animation
4		1810015	Female	40	Movie
5		1810015	Female	46	Animation
6		1810013	Female	43	Drama

k-anonymization

k-anonymized data

No.	Post code	Sex	Age	Hobby
1	18000**	Male	30s	Animation
2	18000**	Male	30s	Animation
3	18000**	Male	30s	Animation
4	18100**	Female	40s	Movie
5	18100**	Female	40s	Animation
6	18100**	Female	40s	Drama

Fig. 6. Example of k-anonymization (k=3).

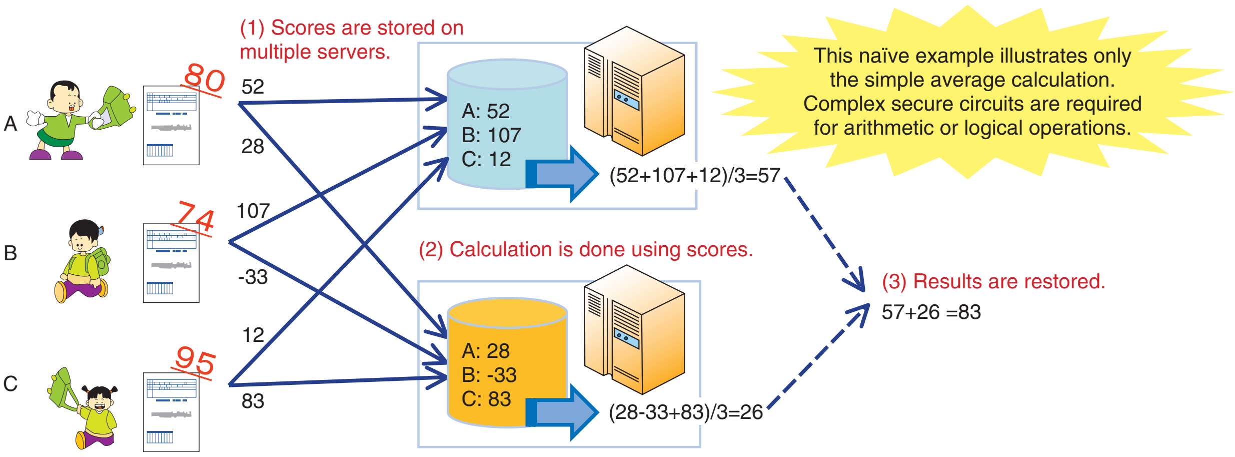

(1) Scores are stored on multiple servers.
A
B
C
80
74
95
52
28
107
-33
12
83
A: 52
B: 107
C: 12
(52+107+12)/3=57
A: 28
B: -33
C: 83
(28-33+83)/3=26
(2) Calculation is done using scores.
(3) Results are restored.
57+26 =83
This naïve example illustrates only the simple average calculation. Complex secure circuits are required for arithmetic or logical operations.

Fig. 7. Simple example of average score calculation using SFE.

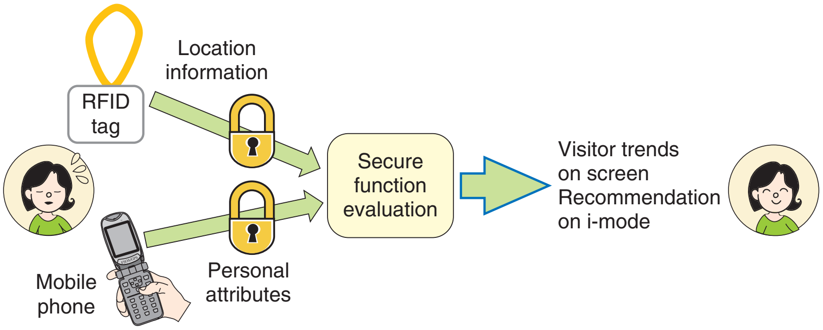


Fig. 8. SFE demonstration experiment.

shown in Fig. 7, as the basis of data protection. For the processing of sensitive data like lifelogs, a technique like SFE is needed and a cloud service that deploys such secure functions is desirable. To date, SFE has been generally believed to be impractical because of its slowness, but NTT's SFE has demonstrated the ability to calculate the maximal value of 1000 data items within one second, which proves its practicality.

6. SFE demonstration experiment

NTT Information Sharing Platform Laboratories conducted an SFE demonstration experiment at NTT R&D Forum in February 2010 (**Fig. 8**). In this experiment, we succeeded in analyzing the location information of forum participants while keeping personal information secret. 1000 visitors wore radio-frequency identification (RFID) tags to record their locations. In addition, they registered their personal attributes, such as affiliations and occupations, from personal computers or mobile phones. The collected information was analyzed safely using NTT's SFE to show the trend of visitors to each exhibition and to recommend exhibits for each visitor [4].

7. Conclusion

This article mentioned legal issues for lifelog utilization, introduced privacy preserving data utilization (PPDU) technologies, and described research and development in NTT Laboratories. We are working hard to make PPDU practical and to provide a platform for sharing lifelogs safely with assured privacy.

References

[1] METI Information Grand Voyage.
http://www.meti.go.jp/policy/it_policy/daikoukai/igvp/cp_en/index.html

[2] MIC (in Japanese).
http://www.soumu.go.jp/menu_news/s-news/02kiban08_02000041.html

[3] K. Takahashi, K. Hirota, K. Chida, and D. Ikarashi, "A Study of Privacy Preserving Data Utilization," Computer Security Symposium (CSS) 2009, Information Processing Society of Japan, Toyama, Japan (in Japanese).

[4] K. Shibata, K. Chida, D. Ikarashi, T. Yamamoto, and K. Takahahshi, "Delegated Two-party Secure Function Evaluation System with Spreadsheet Front-end," Computer Security Symposium (CSS) 2009, Information Processing Society of Japan, Toyama, Japan (in Japanese).

Katsumi Takahashi

Senior Research Engineer, Supervisor, Group Leader of Information Security Project, NTT Information Sharing Platform Laboratories.

He received the B.S. degree in mathematics from Tokyo Institute of Technology and the Ph.D. degree in information science and technology from the University of Tokyo in 1988 and 2006, respectively. He joined NTT Laboratories in 1988 and has studied information retrieval, data mining, location information processing, information security sociology, privacy preserving techniques, and cryptographic techniques. He has developed several commercial systems including i-Townpage, Mobile Info Search, and privango.