Technologies for Improving Reliability of Personal Data Distribution Platforms to Realize Data-Driven Society

● Yuji Yamaoka ● Takao Ogura ● Hidenobu Oguri ● Kouichi Ito

As expectations for data utilization are rising around the world, the Japanese Government is promoting information banks, which are businesses that facilitate the secure distribution and utilization of personal data, expressly with the individual's involvement. To ensure the reliability of these information banks, the government is encouraging the private sector to establish a certification system. In response, the private sector has launched projects to review and certify information bank businesses. In view of this worldwide trend, Fujitsu Laboratories has developed privacy risk assessment technology that makes it possible to manage data by quantifying data leakage risks, and consent management technology that is capable of aggregating and managing individual consent for distributed personal data. These technologies allow for secure distribution of personal data and contribute to the realization of a data-driven society that gives consideration to privacy. This paper provides overviews of the two developed technologies, example applications of privacy risk assessment technology, and a performance evaluation of consent management technology.

1. Introduction

As expectations for data utilization are rising around the world, the Japanese government has established concepts such as Society 5.0[1)] to expressly encourage the distribution of personal data. For example, the Ministry of Internal Affairs and Communications is promoting information banks, which are businesses that promote the distribution and utilization of personal data based on the involvement of the corresponding individual. To ensure the reliability of these information banks, the government is encouraging the private sector to establish a certification system. In response, the Information Technology Federation of Japan started reviewing and certifying information bank businesses in December 2018.[2)]

Fujitsu provides Hybrid IT and Cloud Services[3)] as infrastructure for building information banks. Furthermore, it is also engaged in the practical application of the distributed personal data store (PDS) technology that drives the Personium service, and other activities. In recognition of the contributions that these activities have made to the development of the industry, Fujitsu received an award at the 66th Electrical Science and Engineering Promotion Awards.

Information banks are typically exposed to high operational risks because they distribute personal data. In addition to measures for ensuring the safe management of information, those for preventing leakage from data utilizers must be required. Also, under the certification system, information banks must assume responsibility for compensating individuals for damages arising from accidents such as leakage from data utilizers.

In view of these trends, Fujitsu Laboratories has developed privacy risk assessment technology[4)] that facilitates the personal data leakage risk quantification and management required for the continued stable operation of information banks over the long term. Furthermore, it has also developed consent management technology capable of aggregating and managing consent information for distributed personal data.

This paper provides overviews of the two developed technologies, example applications of privacy risk assessment technology, and a performance evaluation of consent management technology.

2. Privacy risk assessment technology

This section presents an overview of the privacy risk assessment technology and examples of its application.

2.1 Overview of the technology

Fujitsu's privacy risk assessment technology quantifies and expresses as a monetary value the privacy risks when personal data is exposed (**Figure 1**). A model that calculates the identifiability of data after anonymization (degree of low anonymity), which was previously lacking, and the high-speed calculation technology that depends on this model are key. It has been confirmed that when using a computer with typical performance, the technology is sufficiently practical to calculate an actual data set for one million people in approximately one hour.

Previously, there were no quantifiable guidelines for the amount of compensation for damages when personal data leaked from an information bank, which meant that measures for reducing risks were not taken and potential risks remained. This technology resolves this problem.

2.2 Risk assessment examples

This subsection presents two new examples of applying the privacy risk assessment technology, for local government data and medical data.

1) Local government data

An example of applying privacy risk assessments in local government is the data analysis demonstration test held jointly between Otsu City in Shiga Prefecture and Fujitsu.[5)]

After the establishment of the Basic Act on the Advancement of Public and Private Sector Data Utilization, many local governments tried to create a plan to promote the utilization of data. However, because the movement and processing of data even within the local government was restricted, it was not possible to integrate data held in different departments or to freely analyze and discuss the data.

In response to this issue, Otsu City and Fujitsu anonymized the data using FUJITSU Business Application NESTGate Anonymization V1[6)] to enhance data security. This enabled 15 database types relating to childbirth and childcare to be collected and integrated from within the local government and then to be utilized for actual policymaking. However, local government officials cannot determine whether or not data can be used just because it has been anonymized. It is essential that local government administrators who are responsible for explaining their actions to their citizens establish systems for visually verifying that the risks of processed data are sufficiently low, and store this information as evidence.

To do this, various types of anonymized data were assessed using the privacy risk assessment tool (**Figure 2**). In this assessment, the following anonymization was performed for the three personal data items of nationality/gender, address, and date of birth, and the risks were assessed for each data combination.

- Nationality/gender: handling of unknown persons
  Yes: Both nationality and gender. Unknown persons included.
  No nationality: Gender only.
  Neither: Neither nationality nor gender. Unknown persons deleted.
- Address
  Area: 231 areas in Otsu City
  District: 9 area groups defined by the local government

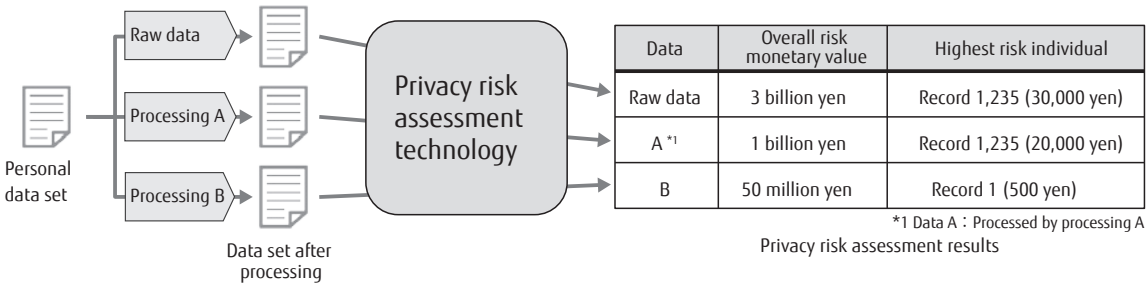


Figure 1
Privacy risk assessment technology.

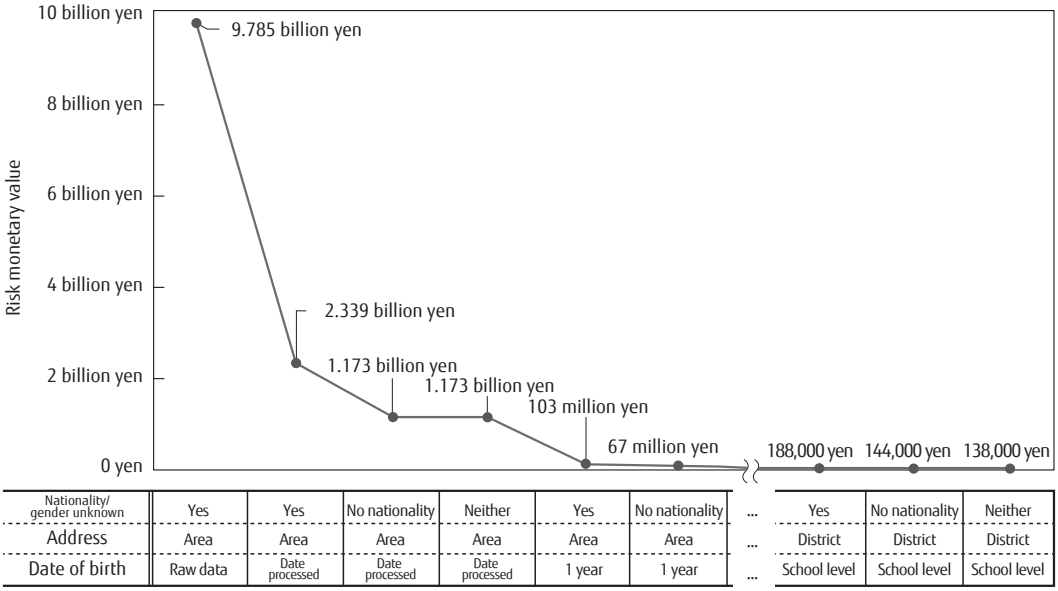


Figure 2
Effects of anonymization in Otsu City.

- Date of birth
  Raw data: Year, month, and day of birth
  Date processed: Year and month of birth
  1 Year: Year and month of birth
  School level: Classified as elementary school/middle school/high school/university

The overall risk value for the raw data before anonymization is approximately 9.8 billion yen. In contrast, the value after the strictest processing (Nationality/ Gender: None, Address: District, Date of birth: School level) is approximately 140,000 yen. This shows specifically that the risk is reduced to 0.001% of the raw data.

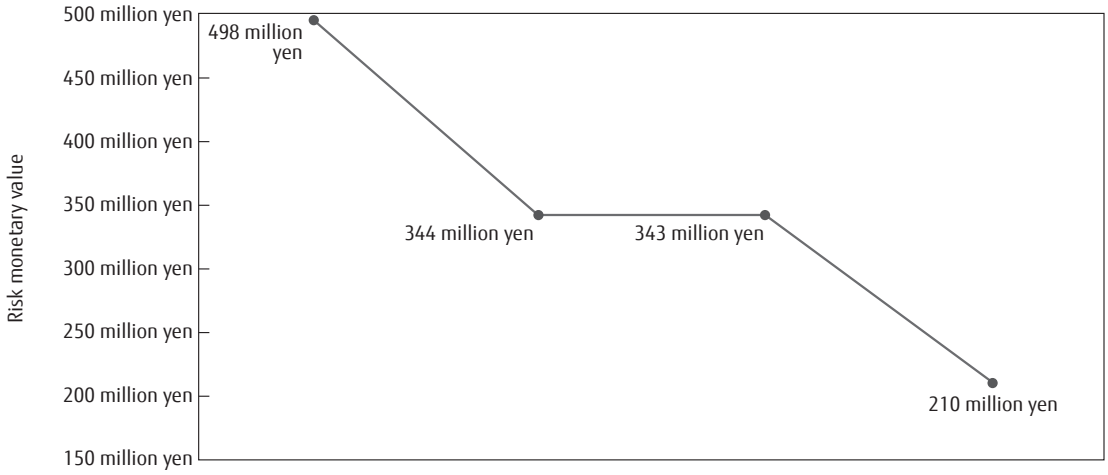
Based on these results, Otsu City established a policy to enable data with a risk of 0.1% or less of the raw data to be used after a simple application is made to the local government office. This quantification of safety also had the effect of sharing the usefulness and hazards of using personal data within the local government. Consequently, local government officials rated the tool as very beneficial in promoting data utilization.

2) Medical data

In Japan, the Cancer Registry Promotion Act was enforced in January 2016. The National Cancer Registry[7] was established under this law as a system for collecting information related to cancer examinations, progress, and other treatments, and to store, manage, and analyze this information. This enabled the National Cancer Center Japan to create a database of cancer-related information that can be used by national and local government organizations. Anonymized information is also provided to researchers who want to conduct research related to cancer. In joint research with the National Cancer Center Japan, Fujitsu assessed the privacy risk assessment technology, assuming the use of anonymized information.

In terms of the risk of identifying an individual by special group information like a cancer register, the relationship between the anonymization level and the privacy risk was assessed using the privacy risk assessment technology for anonymized data that was provided for research purposes by regional cancer registries.[8] In this assessment, the following anonymization processing was performed for the three personal data items of age, address, and examination date, and the risks were assessed for each data combination.

- Age
  10-year groups: Ages converted into 10-year units
  10-year groups and young/elderly groups:
  10-year units and consolidated groups of 20 years old or younger and 90 years old or older
- Address
  Municipality: Municipalities within prefecture
  Medical district: 4 medical districts within prefecture defined by Japanese Government



Age	10-year groups	10-year groups	10-year groups and young/elderly groups	10-year groups
Address	Municipality	Municipality	Municipality	Medical district
Examination date	By day	By year	By year	By year

Figure 3
Cancer registration risk assessment results (when sensitivity of disease name is set high).

- Examination date
  By day: Year, month, and day of examination (Raw data)
  By year: Year of examination

It was confirmed that when the strictest anonymization process was applied, the privacy risk could be reduced to 48% of the raw data (**Figure 3**).

As a result, this technology was assessed to be capable of providing certain criteria for personal identification risks that could be judged and determined by ordinary people, even for the anonymized information of cancer registry users.

3. Consent management technology

This section presents an overview of the consent management technology and an evaluation of performance when using this technology.

3.1 Overview of the technology

Fujitsu's consent management technology safely distributes personal data for which distributed management between multiple data providers is performed, while aggregating and managing information related to an individual's consent.

Information banks are assumed to function by distributing data via aggregation and management on a central server. However, personal data is often managed outside this framework by other data providers. For example, personal data such as driving records are updated frequently, therefore distribution without aggregation is preferable. However, the consent of the individual is usually required to utilize personal data, therefore it is better to efficiently aggregate the results of consent.

This creates the problem of using different management systems according to the data type. To resolve this issue, the User Managed Access (UMA)[9] protocol was extended to enable the efficient distribution of personal data for multiple people.

1) Consent portal

The consent portal is a system for obtaining the consent of individuals whose data are stored at the data provider on behalf of the data utilizer, when personal data is supplied by the data provider to the data utilizers. The portal also manages the records of such transactions (**Figure 4**). The system enables individuals to reference information such as past consent records, vendors who have provided data, and the contents of personal data already provided. Furthermore, the system allows different vendors to use the same

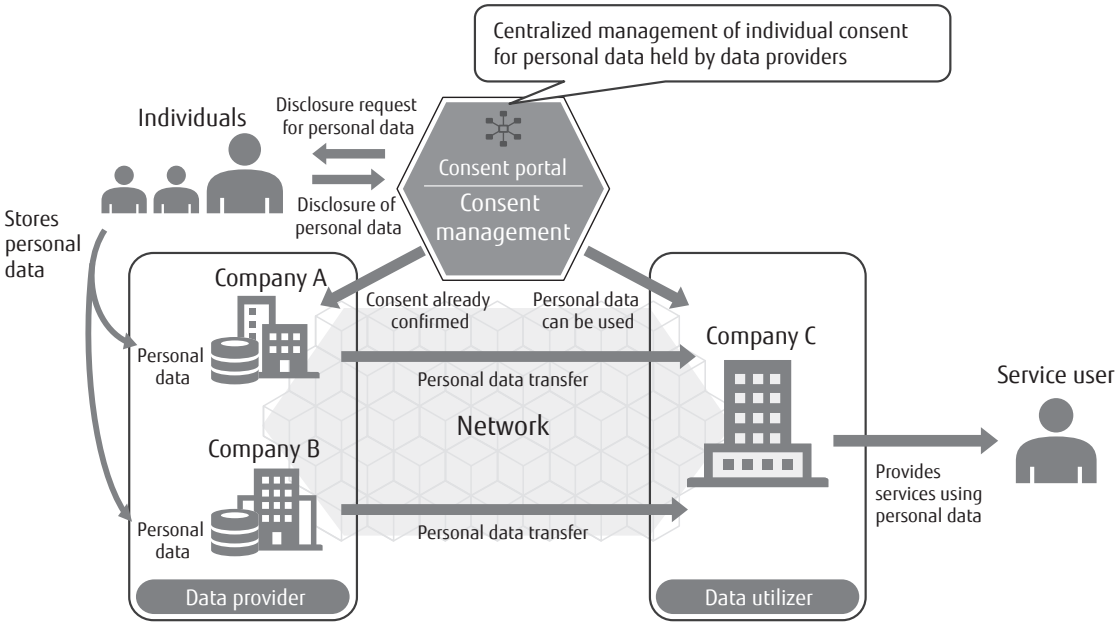


Figure 4
Consent portal.

terms of service and consent computer screens, and the differences between the terms of service and consent contents that individuals have agreed to in the past can be clarified, which makes it easy to determine whether they have given consent.

By using the consent portal, data utilizers who provide services can reduce the costs involved in independently developing terms of service, text and computer screens for consent, and consent record systems. For example, in the case of a telematics insurance service, the insurance company (the data utilizer) acquires the driving-record data of individuals from the automobile company (the data provider) in order to calculate insurance premiums. In this service, the consent portal becomes a contact point for individuals, which makes it easy to obtain consent.

2) Method for batch acquisition of personal data

To enable the analysis of personal data based on attributes such as age or gender, technology was developed for the batch acquisition of personal data with the same attributes. The UMA protocol standard enables personal data to be shared between different organizations such as data providers and data utilizers, and for access to be controlled from a single location. UMA was designed to acquire data on the basis of individual users. This causes problems, for example, when acquiring large-scale personal data for 10,000 people, because a large number of messages are generated.

In response, Fujitsu Laboratories developed a function for the batch acquisition of data with the same attributes using a single UMA protocol sequence. The consent portal handles grouped data for multiple users who have given their consent virtually as the data of a single person, and sets access rights as a batch to this data group. This enables the data utilizers to acquire personal data in batches.

Also, by describing the attribute conditions (such as age or gender) of the data to be acquired in the Unified Resource Identifier (URI) of the data request, consideration can be given to the compatibility of the interface on the side of the data utilizers in the UMA protocol. This enables the data utilizers to use UMA to acquire data both on an individual basis, and by batch. For example, an open source format that supports UMA can be used to develop various services that utilize data.

3.2 Performance evaluation

To evaluate the consent management technology, a prototype system was built based on a telematics insurance service and the performance of the personal data batch acquisition method was evaluated. This

service assumed that user consent had been acquired in advance at the time of subscription to enable the insurance company to use the driving records of the vehicle.

In terms of the performance evaluation environment, a consent portal was built on a single server[note)] that could be used by multiple data providers, and a measurement tool was built on a separate server. The HTTP request for acquiring data was sent from the measurement tool to the target server, and the response time was measured.

In the performance evaluation, requests were sent from the simulated measurement tool to the data utilizer with the following three methods, and the time required to acquire the personal data of 200 people was measured. The size of the personal data was changed between 10 KB, 100 KB, and 1 MB.

- Batch acquisition method: Personal data for 200 people is acquired with a single request.
- Sequential method: Request is sent from the measurement tool, and the next request is sent after the response is received. This is repeated 200 times.
- Parallel method: 200 requests are sent from the measurement tool every 10 milliseconds.

The results of the performance evaluation showed that the batch acquisition method was faster by 6 to 236 times compared to the sequential method and parallel method (**Figure 5**).

Based on these results, the time was calculated for the batch acquisition of the driving-record data for 10,000 people (assuming approximately 100 KB per person). Assuming a transmission speed of 10 Gbps, the UMA protocol process takes approximately 800 milliseconds (test value) and the personal data transfer takes approximately 3 seconds (calculated from a transmission speed of 10 Gbps). Therefore, the data for 10,000 people can be acquired in approximately 3.8 seconds.

As a result, the UMA extension enables the efficient distribution of personal data for a large number of people.

4. Conclusion

This paper described the privacy risk assessment technology and consent management technology developed by Fujitsu Laboratories in order to increase the reliability of personal data distribution businesses such as information banks. These technologies not only make it easier for individuals to consent to the distribution of personal data, but they also enable new types of data analysis to be performed by distributors. Such systems can also improve the accuracy of analysis, which will contribute to the realization of a data-driven society.

In the future, the practical application of these technologies will increase the reliability of personal data distribution businesses and expand the distribution and utilization of personal data, which will hopefully grow the economy and help solve social issues.

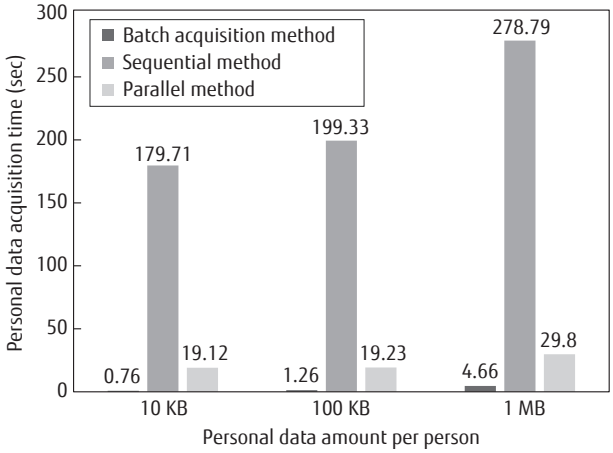


Figure 5
Performance evaluation of consent management technology.

note) CPU: Intel E5-2660 2.60 GHz, 10 cores, RAM: 64 GB.

References

1) Japanese Cabinet Office: Society 5.0. *https://www8.cao.go.jp/cstp/english/society5_0/index.html*
2) Information Technology Federation of Japan: Information Bank Promotion Committee. (in Japanese). *https://www.tpdms.jp/*
3) Fujitsu: Hybrid IT and Cloud Services. *https://www.fujitsu.com/global/services/hybrid-cloud/*

4) Y. Yamaoka: Digital Security Systems Protecting Society from Potential Threats. FUJITSU Sci. Tech. J., Vol. 54, No. 5, pp. 56-61 (2018).
https://www.fujitsu.com/global/documents/about/resources/publications/fstj/archives/vol54-5/paper08.pdf
5) Otsu City, Fujitsu: Otsu City and Fujitsu Agree to Collaboration in Fields of ICT Utilization and Data Analysis. (in Japanese).
https://pr.fujitsu.com/jp/news/2018/11/8.html
6) Fujitsu: FUJITSU Business Application NESTGate Anonymization V1. (in Japanese).
https://www.fujitsu.com/jp/solutions/business-technology/intelligent-data-services/bigdata/ba-solutions/nestgate/anony/index.html
7) National Cancer Center Japan: Provision of National Cancer Registry, Registration Information. (in Japanese).
https://ganjoho.jp/reg_stat/can_reg/national/datause/general.html
8) Japan Association of Cancer Registries: About cancer registration. (in Japanese).
http://www.jacr.info/about/registry.html
9) Kantara Initiative: User Managed Access.
https://kantarainitiative.org/confluence/display/uma/

Yuji Yamaoka
Fujitsu Laboratories Ltd.
Mr. Yamaoka is currently engaged in the research and development of technologies for data and privacy protection.

Takao Ogura
Fujitsu Laboratories Ltd.
Mr. Ogura is currently engaged in the research and development of technologies for data distribution and utilization.

Hidenobu Oguri
Fujitsu Laboratories Ltd.
Dr. Oguri is currently engaged in the research and development of technologies for data and privacy protection.

Kouichi Ito
Fujitsu Laboratories Ltd.
Dr. Ito is currently engaged in the research and development of technologies for data protection in Europe.