Key Points of the Amendments to the Act on the Protection of Personal Information, and Anonymization Methods for the Use of Personal Data

Kumiko Kameishi, Keiichi Hirota, Akiko Fujimura, Fumihiko Magata, and Yukiyoshi Ota

Abstract

The Act on the Protection of Personal Information was amended in 2015 to promote further development of industry through the use of personal data while at the same time protecting people's privacy. In this article, we discuss the five amendments that were made to this act. We also describe an anonymously processed information system that could lead to the creation of new business for NTT, and we introduce NTT's proprietary Pk-anonymization technology that keeps information secure without harming its usefulness.

Keywords: Amended Act on the Protection of Personal Information, anonymization methods, personal data

1. Introduction

Amid calls for rules governing the use of big data and personal data that take legal issues and privacy into consideration, the Act on the Protection of Personal Information was amended in 2015 (referred to as the *amended law*[*1] [1] below), and government-led preparations are now being made prior to the full enforcement of this amendment on May 30, 2017. The structure of laws and ordinances relating to personal information is shown in **Fig. 1**. At the lowermost part of this structure, the Act on the Protection of Personal Information provides the basis for higher-level rules and guidelines, and the following items in the upper half of the structure should be seen by those trusted with personal information: Guidelines on Personal Information Protection Law [2], ministry guidelines, and Guidelines for Accredited Personal Information Protection Organizations.

Although the Guidelines for Accredited Personal Information Protection Organizations originally received a quiet reception, the new legal reforms are particularly important because they add items relating to the handling of *anonymously processed information* and make it obligatory for organizations to provide guidance and also include recommendations for compliance with the guidelines. The accredited personal information protection organizations[*2] that apply to the NTT Group include the Japan Data Communications Association (JADAC) and the Japan Institute for Promotion of Digital Economy and Community (JIPDEC).

*1 Amended law: The partial amendments on the Act on the Protection of Personal Information and the Act on the Use of Numbers to Identify Specific Individuals in the Administrative Procedure.

*2 Accredited personal information protection organization: A private organization that handles complaints and provides information for businesses for purposes such as ensuring the correct handling of personal information. There are currently 42 such organizations in diverse fields. Each organization formulates and publishes policies based on related ministry guidelines.

*Parts outlined with a dotted line have yet to be published (as of 19th April 2017).

Fig. 1. Structure of the Amended Act on the Protection of Personal Information.

2. Key amendments to the Act on the Protection of Personal Information

We will refer to **Fig. 2**, which was prepared based on government materials relating to the key amendments to the Act on the Protection of Personal Information, to describe five points that appear to have a particularly large impact on business, based on the contents of the promulgated enforcement rules and regulations.

2.1 Clarification of the definition of personal information: introduction of individual identification codes (amended law, Article 2, Paragraph 2)

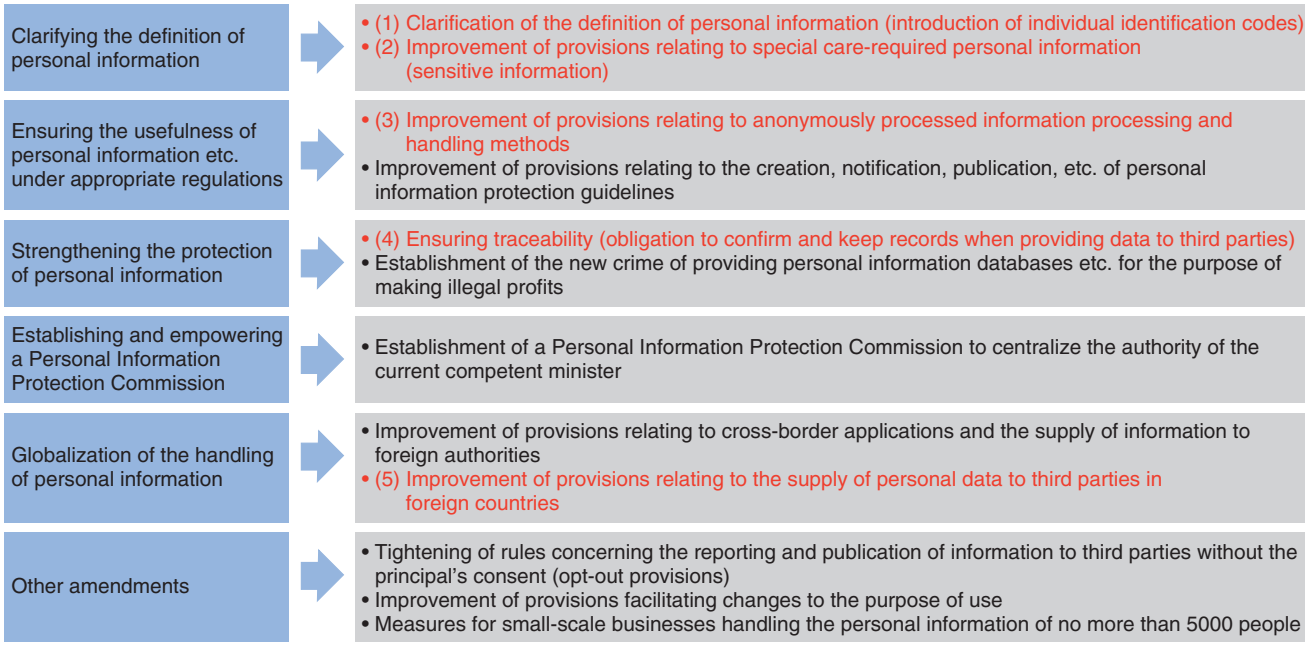
The concept of individual identification codes was introduced as a way of identifying specific individuals in a set of information so that personal information could be referenced without using a person's name or other details. Specifically, it includes biometric information such as fingerprint/face authentication data and vein pattern data, *My Numbers* (social security/tax numbers), passport numbers, driving license numbers, and pension account numbers. For example, an individual set of fingerprint data for authentication stored inside a smartphone or USB (universal serial bus) memory stick with a fingerprint authentication function constitutes personal information, thus making it necessary to check the provisions regarding how this information is handled by businesses.

2.2 Personal information requiring special care (amended law, Article 2, Paragraph 3)

Personal information that must be handled carefully typically comprises information that would be liable to cause discrimination and/or prejudice if mishandled. This includes sensitive information such as the principal's race, creed, social status, and medical history. Since the introduction of this amendment, personal information that must be handled carefully can no longer be handled without consent, so it is necessary to obtain in advance a principal's consent even for simple workplace health questionnaires and the like.

2.3 Anonymously processed information (amended law, Articles 36–39)

The amended law defines anonymously processed information as information that has been processed to make personal information impossible to identify a specific individual, and from which it is impossible to restore this person's personal information, and a system where this information can be distributed subject to certain regulations. When supplying personal information to a third party, it is necessary to obtain the principal's consent, but anonymously processed information has the advantage that the principal's consent does not have to be obtained.



Source: Website of Cabinet Secretariat, bills to the 189th ordinary Diet session, Mar. 2016 (in Japanese).
http://www.cas.go.jp/jp/houan/150310/siryou1.pdf

Fig. 2. Key amendments to the Act on the Protection of Personal Information.

2.4 Ensuring traceability (obligation to confirm and keep records when providing data to third parties) (amended law, Articles 25 and 26)

A service provider that receives personal data from another party is subject to various obligations including confirming the background of what was transferred, and keeping records including the items of information that were transferred, and when the transfer took place. This is intended to ensure that personal information obtained and divulged by illegal means is prevented from circulating endlessly. This requires checking whether an existing contract has been commissioned or has third-party provisions, and if necessary adding log acquisition functions.

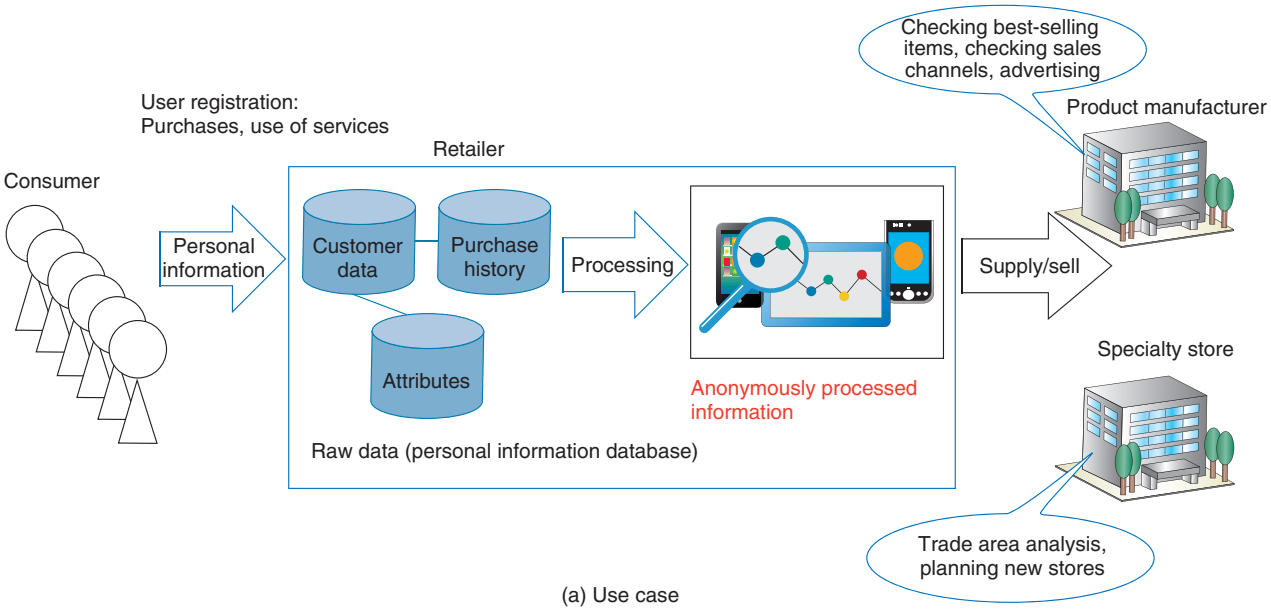
2.5 Provision of personal data to third parties in foreign countries (amended law, Article 24)

A third party in a foreign country is defined as a person or persons (including corporate bodies) located in a foreign country (but not a company's foreign branch offices or the like). When personal data are provided to a third party in a foreign country, it is in principle not possible to do so without the principal's consent. This consent must be obtained even when entrusting employee information to a cloud service provided by a foreign entity. Exceptions to this rule include businesses in countries designated by the Personal Information Protection Commission, and businesses that have been certified based on the APEC Cross Border Privacy Rules (CBPR) system, for which consent is not required [3].

3. Creation of new business through the use of anonymously processed information

Anonymously processed information is information related to an individual that has been processed to protect people's identities from being disclosed. It has the advantage that the principal's consent is not required when the information is provided to a third party, and it is therefore expected that new business will be created to take advantage of this feature.

Some use cases of anonymously processed information are illustrated in **Fig. 3**. When a consumer registers with a retailer in order to access services or purchase goods, the retailer stores raw data such as the consumer's customer data and purchase history. The retailer might choose to process these raw data and supply or distribute the processed data so that other people can use them to identify best-selling items, analyze sales channels, or plan the opening of



(a) Use case

(i) Raw data (personal information)

Customer ID	Registration date (YYYYMM DD)	Name	Address (full address including postal code)	Date of birth (YYYYMM DD)	Gender (male/ female)	Purchase history (date, purchase details)	Where purchased (1: In store, 2: Online)	Payment method (1: Cash, 2: Credit card)	Attribute 1 (home delivery requested)	Attribute ...

Processing

- Delete or replace items that could identify individual customers (names, etc.).
- Round off numerical values, etc.

(ii) Anonymously processed data

Provisional ID	Address (partial address, specifying city and district only)	Age bracket (in 10-year increments)	Gender (male/ female)	Purchase history (year and month, category)	Where purchased (1: In store, 2: Online)	Attribute 1 (home delivery requested)	Attribute ...

(iii) Statistical information

Product category	Purchase amount by age bracket (e.g., monthly sales trends)							
	10s	20s	30s	40s	50s	60s	70 and over	Total amount
Food								
Commodity								

(b) Processing the data set

Fig. 3. Conceptual illustration of anonymously processed information usage.

Table 1. Examples of anonymously processed information methods.

Method	Overview
Deleting items/records/cells	Delete descriptions of personal information included in personal information databases and the like before they are processed.
Generalization	Descriptions included in the information to be processed can be replaced with higher-order concepts or numerical values, and numbers can be rounded to the nearest whole value. For example, *cucumbers* could be replaced with *vegetables* in purchase history data.
Top (bottom) coding	This is the process of especially large values and especially small values in the numerical values being summarized and integrated into a personal information database requiring anonymization. For example, in age-related data, the numerical data for people aged 80 and over should be summarized as "≥80 years" data.
Micro-aggregation	After personal information consisting of a personal information database or the like is grouped together in preparation for anonymization processing, it should be replaced by a representative description of the group.
Data exchange (swap)	The process of anonymizing a personal information database or the like by (randomly) swapping the descriptions and other information included in the personal information constituting the database
Addition of noise (errors)	The addition of random numbers with a certain distribution so numbers can be replaced with other arbitrary values
Pseudo-data generation	The creation of artificially synthesized data and including the data in a personal information database during anonymization processing

new stores. It is possible to use data anonymization for this subject.

Possible ways of processing these data are shown in Fig. 3(b). In table (i), the retail operator manages users according to their customer identifications (IDs) and records each customer's registration date, name, address, date of birth, and gender. Furthermore, each customer ID is associated with information including a purchase history (date, purchase details, amount paid), where purchased (in store, online), payment method (cash, credit card), and whether or not a home delivery service was used. Under the current law, this information might have been provided to product manufacturers in the form of statistical information as shown in table (iii), but one might consider processing these data in such a way that they contain slightly more detail without revealing any personal information.

As in table (ii) above, the information is processed so that it can be used to find out how many products in a particular category were purchased during a particular period, and whether they were purchased in store or online. To avoid identifying individuals, the number of purchases is rounded, and any information that could be used to identify someone (such as their name or address) is deleted. The processed data are expected to be used for marketing purposes such as allowing product manufacturers to check the performance of strong sellers, or finding out if a product category of interest has been accepted into the envisaged customer layer, or whether or not home deliveries are popular.

4. Anonymization methods we have developed

In anonymization processes according to the amended law, personal information is required to be processed so as to make it impossible to identify specific individuals. Since it is very difficult to prove that a specific individual cannot be identified, the guidelines [4] state that in practice, rather than requiring the elimination of all technical possibilities of identifying a person by any means whatsoever, it should at least be impossible for a personal-information-handling business operator or an anonymously processed information-handling business operator to identify a specific individual using ordinary business skills and methods.

Specific methods and standards for anonymization processing are to be determined separately for each industry according to the abovementioned detailed regulations and guidelines. For example, in the anonymously processed information guidelines [4] and the Anonymously Processed Information Creation Manual [5], methods such as top (bottom) coding and noise addition are presented (**Table 1**). In practice, when these methods are used for anonymization processing, it is necessary to study what sort of processing methods should be applied by clarifying the use cases of personal data, and identifying the risks of outcomes such as the identification of individuals by partitioning the data items according to identifiers, attributes, and history.

To use these data for actual business, in addition to ensuring that they are anonymized and do not allow

the identification of individuals, there is also a greater need to process the data so that they can be used effectively. At NTT Secure Platform Laboratories, we have developed an anonymization method based on NTT's own evaluation measure called Pk-anonymity where the evaluation measure of k-anonymity is replaced with a probabilistic measure [6]. This is the first ever method that introduces randomness by stochastic rewriting of item values, which has been mathematically proven to have a level of security equivalent to that of k-anonymity. In anonymization based on Pk-anonymity, the values are stochastically rewritten and are thus different from the original data, but can be processed into data that are statistically close to the original data by using a probability-based method called Bayesian inference. A processing method called generalization is often used in k-anonymization. For example, the word *cucumbers* might be changed to *vegetables* in order to change the level of detail in the data. However, in Pk-anonymization, the data are rewritten without changing the level of detail. For example, *cucumbers* might be changed to *tomatoes*. It is thought that this method could be useful for the analysis of marketing data or the like where people want to see detailed data distributions.

At NTT Secure Platform Laboratories, we are preparing to provide services to tie in with the enforcement of the amended law by continuing to research and develop anonymization methods such as these, by participating in joint research initiatives and verification trials together with businesses that actually use this sort of data, and by taking part in national contests related to anonymization methods in order to gather technical know-how related to the anonymization of diverse types of data.

5. Future prospects

As progress is made with the related guidelines of the amended law and revisions to the Guidelines for Accredited Personal Information Protection Organizations, we also plan to review the provisions related to the protection of personal information by businesses, and to investigate business models that make effective use of anonymized data. In line with current trends in legal systems, we will continue to support the NTT Group in both legal and technical aspects and continue to take part in external activities including academic activities.

References

[1] T. Hioki and Y. Itakura, "Mechanism of the 2015 Amendment to the Act on the Protection of Personal Information," Shojihomu, 2015 (in Japanese).
[2] Personal Information Protection Commission JAPAN, http://www.ppc.go.jp/en/
[3] Personal Information Protection Commission, "Guidelines on Personal Information Protection Law (Provision to a Third Party in a Foreign Country)," 2016 (in Japanese).
[4] Personal Information Protection Commission, "Guidelines on Personal Information Protection Law (Anonymously Processed Information)," 2016 (in Japanese).
[5] Ministry of Economy, Trade and Industry, "Reference Material for Use by Service Providers when Considering Methods for the Creation of Anonymously Processed Information (Anonymously Processed Information Creation Manual) Ver. 1.0," 2016.
[6] "Focus on the News: Development of a New Personal Data Anonymization System for the Big Data Era—Providing Advanced Privacy Protection While Retaining the Data's High Utility Value," NTT Technical Journal, Vol. 26, No. 5, pp. 51–52, 2014 (in Japanese).

Kumiko Kameishi
Senior Research Engineer, Secure Architecture Project, NTT Secure Platform Laboratories.
She received a B.S. and a Master of Environmental Science from University of Tsukuba in 1989 and 1991. She jointed NTT Telecommunication Networks Laboratory in 1991. She is studying information security and privacy issues of personal data services. She is a member of the Information Processing Society of Japan (IPSJ).

Keiichi Hirota
Senior Research Engineer, Data Security Project, NTT Secure Platform Laboratories.
He received a B.S. and M.S. from Mie University in 1995 and 1997, and a Ph.D. in informatics from the Graduate University for Advanced Studies (SOKENDAI), Kanagawa, in 2008. He joined NTT in 1997. His current research interests include security and privacy in information processing, information sharing, and data utilization. He is a member of IPSJ.

Akiko Fujimura
Research Engineer, NTT Secure Platform Laboratories.
She received an LL.B. and a Master of Media and Governance from Keio University, Kanagawa, in 1997 and 1999, and a J.D. from Chuo University, Tokyo, in 2008. She joined NTT Information Sharing Platform Laboratories in 1999 and has been engaged in research on technological and legal issues of information security, personal information protection, and privacy protection.

Fumihiko Magata
Senior Research Engineer, Secure Architecture Project, NTT Secure Platform Laboratories.
He received an LL.B. from Chuo University, Tokyo, in 1992. He jointed NTT in 1992. He is currently studying information security in the interdisciplinary field of social science and information engineering. He is a member of the Japan Society of Security Management and the Information Network Law Association. He is a Professional Engineer (Information Engineering).

Yukiyoshi Ota
Senior Research Engineer, Secure Architecture Project, NTT Secure Platform Laboratories.
He received an M.E. in electrical engineering from Osaka University in 1992 and jointed NTT the same year. He is currently researching information security. He is a member of the Institute of Electronics, Information and Communication Engineers.