The Concept of Personal Information for Data Consumption, Use, and Protection

Ji-Yeon YOO
Associate Professor
Department of Information and Security Management
Sangmyung University
Seoul City 03016
Republic of Korea

Abstract

With advancements in information technology and information services, personal information is being combined through various channels like smartphones, Social Network Service, and the Internet of Things. Personal information is in a gray zone where distinguishing levels of identity and combination is difficult. The question that arises is whether "information uploaded by a person" or "personal data" can be considered as personal information. Therefore, we need to discuss and agree on expanding the concept of personal information from that focused on personal identification to "personally related information," which includes personal data that can estimate an individual's personal state. We must simultaneously discuss how far "personally related information" should be regulated if "personal information" is expanded to include it. In other words, the question comes down to how to measure the sensitivity of personal information. For this, discussions on personal identity are examined and standards to classify personal information are located.

Keywords: Personal Information Protection; Personal Information Definition; Personally Identifiable Information (PII); Non-Personally Identifiable Information (Non-PII); Personal Data

1. Existing Definition of Personal Information and Discussion of Concerns

1.1. Definition of Personal Information

So far, personal information has been classified into "personally identifiable information" (PII) and "non-personally identifiable information" (non-PII), based on personal identity. PII that specifies an individual indicates 1) information that identifies a person, 2) information that can identify a person or find a location, and 3) information that can extract a personal identifier or contact information. "Personal identifier" includes name, address, phone number, fax number, email address, financial information, medical history, social security number, and credit card information, and any other information that can identify a specific individual. If a personal profile, unique identifier, biometrics, and an IP address are connected with PII, the information is also considered PII. In contrast, non-PII indicates anonymous information that cannot specify an individual (FTC, 2000a).

According to the definition under the present legal system, personal information is that which can identify an individual by using name, residential ID, and image; this can identify a specific individual along with other information, although it cannot identify an individual with only that information (Article 2, Clause 1 of the Personal Information Protection Act; Article 2, Clause 6 of the Act on Promotion of Information and Communications Network Use and Information Protection). Thus, the criteria for personal information are based on "identity" and "combination." This definition's classification of personal information has recently become more ambiguous due to "combination." Previously, available personal certification information (e.g., name, residential ID, and social security number) was referred to as PII; all other information was deemed non-PII. However, with advanced information technology and information service, information given to an individual randomly (e.g., through cookies, terminal identification numbers) and unique personal information created by an individual (e.g., Google ID) are produced and activated through their use.

With enhanced data analysis skills and easier individual identification, some opinions and policies suggest that personal information should be treated as PII.

As information produced through personal online activities (e.g., purchase history) or personally produced information (e.g., life-logging information) are accumulated, shared, and reused, violation of privacy is being questioned.

1.2. Discussions on Non-PII

1.2.1. Cookie

"Cookie gate" refers to the case in which companies such as Google collected personal information by using cookies, a file that remembers Internet access information. Google used Internet Explorer and Safari, Internet access programs, to obtain personal information by pretending that users requested the security system's temporary release when they clicked on advertisements. Google still tracks users' personal history through the use of cookies—even when users logout of their accounts—and collects and saves the information. If users login again, personal information collected in the logout state is automatically added to a previously accumulated personal information file. Using cookies, Google has unauthorized access to personal information saved in users' computers? Although cookie information is non-PII, significant concern is expressed regarding cookies containing identity in combination with profiling. Therefore, review of cookie information is necessary. Beginning on May 25, 2011, the European Union (EU) emphasized the necessity of business operators having explicit agreement from users prior to saving cookies by using the revised European e-Privacy directive, the "EU Cookie Law," and tightening measures concerned.

The British data protection supervisory authority Information Commissioner's Office (ICO) announced that it will apply regulations on prior agreement for website cookies to the privacy directive in the personal information process and telecommunications sector (2002/58/EC) beginning May 26, 2011. The ICO provided one year as a preparatory period for website owners to observe the law. The website had only to notify users of the cookie's use and inform them of how to cancel before revision if they wanted; however, after revision, the website cookie was available only when users agreed to its use. The law requires obtaining agreement on cookie use, including action procedures such as confirmation of cookies in use, their operational procedures, and selection of solutions, pop-ups, and contractual questions (Aboutmyarea, 2011). However, these regulations are rather insufficient as standards for explicit consent of users and user control. Accordingly, the International Chamber of Commerce (ICC) developed the "ICC UK Cookie Guide," which put regulations that require prior consent for practically using cookies, based on the EU's e-Privacy directive (ICC UK, 2012). The ICC UK Cookie Guide classified categories and procedures for cookies that require user agreement into four types:

- *Strictly necessary cookies:* Cookies related to users' e-payment services are prohibited to be used for marketing purposes and do not require user consent when inevitable.
- *Performance cookies*: Cookies used for web improvements like web analysis, response speed, and error management are non-identifiable data, but they require user consent.
- *Functionality cookies:* Cookies controlling user configurations like website individualization, page layout, and storage of user ID require consent to be used for advertisement.
- *Targeting cookies or advertising cookies:* Cookies that collect the majority of user information explicitly require user consent to be used for advertisement.

The ICC Cookie Guide also considered measures for user consent, basically to stop third parties from using cookies. The user consent to cookies should be clearly prepared for each cookie type and included in the user agreement or suggested as a pop-up message to acquire consent when users visit the site for the first time or when the user setting is changed. As not everyone has the same level of technical understanding, a phased self-guide method is available as well (ICC UK, 2012; Global Regulatory Enforcement Law Blog, 2012; BBC news, 2012). The Netherlands also passed the Dutch Telecommunication Act on June 28, 2011, which prohibits unclear consent when using cookies and permits use of cookies only when there is clear prior consent (Privacyassociation, 2011). The United States strongly recommends companies install a "do not track" button, so the user can directly Opt-out with one click in a web browser. This enables users to escape easily from companies' collection of cookie information.

1.2.2. ID

Most Internet services require account registration at the time of use, and login with user ID is obtained. The service couples many actions with one person using the ID. These ID characteristics are also used in identifying personal information.

One example is Cogle, a site that collects personal information fragmentized on the Internet by using various search engine APIs, including Google; Cogle can search for name, email, ID, blog website address, and IP address. Therefore, an ID can potentially connect information to an individual though it might not specify the person. The integrated Google ID ensures that every piece of personal information like searches, social network service (SNS), and location-based behaviors using a smartphone is for only one person.

1.2.3. Profiling

Profiling means developing and using a profile. The dictionary defines profile (as a verb) as "to express features of a person in a systematic way" (*Oxford Dictionary*). However, this meaning has recently been expanded and now indicates collecting information on a user group for effective advertising. What is problematic about profiling is that it combines anonymous non-PII-like advertising cookies with personal information by profile tracking. An advertisement network, in particular, attempts to predict the taste, desire, and purchase habits of an individual consumer and establishes a detailed personal profile by drawing interests and preferences from individual consumers to provide them with advertising targeting specific interests. Meanwhile, consumers never know if their actions are monitored online unless they are notified of the existence of the advertisement network and data collection by the website they visited (FTC, 2000b).

In this regard, the Network Advertising Initiative (NAI), an interest group of online advertisers in the United States, made voluntary covenants in 2000: 1) basically, profiling non-PII of the information object is allowed, but an Opt-out opportunity should be provided to the information object; 2) in the future, PII and non-PII profilers should explicitly notify the user at the time of profiling, and users should be provided an Opt-out opportunity (robust Opt-out); and 3) in case of incorporating profiling of previously collected non-PII into PII, users must consent to collecting any identifiers (Opt-in). Based on this discussion, the standard between PII and non-PII must be reconsidered and the use of personal information for non-PII should be recognized by an individual through Opt-in, on the condition that even non-PII can provide identity when it is accumulated.

1.3. Classification of Personal Information

Recently, a classification of personal information that considers the contractor's unique ID, an identifier created with cookie technology, and an identifier required for login as PII was suggested for privacy policy in Japan. The Japanese Ministry of Internal Affairs and Communications announced initiatives in August 2012 to enhance smartphone privacy with a stronger relationship with users compared with PCs, and classified and provided user information used in smartphones (Ministry of Internal Affairs and Communications of Japan, 2012) (See Table 1). The smartphone privacy initiative classifies user information into user identifier, third-party information, and information on history of behaviors and user state in the communication service. Initiatives consider the contractor's unique ID (ID created by OS [Android ID]), unique device identifier (UDID), international mobile equipment identity (IMEI), international mobile subscriber identity (IMSI), a media access control (MAC) address, login identifier, and identifier created with cookie technology used in smartphones as identifiers. Most IDs in smartphones have characteristic values that cannot be changed by users and are included in user-identifiable information. Third-party information includes data managed in contact information (the third party's name, phone number, and email address).

In the telecommunications service, information on history of behaviors and user state includes history of communications (call log and content, email content, and history of sending and receiving emails), history of behaviors on websites (history of visit, history of purchase, history of search), social information, history of app use (information accumulated by using an app or app-use log, system-use log), location information, and photographs and videos. As stated above, in Japan, information like cookies and IDs used by an individual or information that can estimate an individual's identify from a device used by an individual or services, as well as PII, such as name and address, are considered PII. Information on the individual state, like the history of communications, history of behaviors, and social information, is also considered personal information.

Table 1: Example of User Information in Smartphones

Classification	Type of Information	Information Included
User Identifier	Contractor information (Name, address, etc.)	Name, residential ID, DOB, address, age, gender, and phone number, and personal credit information, like credit card number
	Identifier for login	Identifier like ID used for login to specify the user in the website that provides different services on the network
	Identifier Created using Cookie Technology	Data recorded temporarily in the PC through the web browser when visiting the website (Number of website visits, history on the site, etc.)
	Contractor's unique ID	ID created by OS (Android ID), Unique Device Identifier (UDID),International Mobile Equipment Identity (IMEI), International Mobile Subscriber Identity (IMSI), MAC address, etc.
3rd Party Information	Data managed in Contact Information	Name, phone number, email address, etc.
Information on History of Behavior &User State in Telecommunications Service	History of Communication	Call log, email content, history of sending and receiving email, etc.
	History of Behaviors on Website	History of behaviors such as history of visits, history of purchases, and history of searches in the user website
	Social Information	Behavioral information recorded and accumulated in SNS
	History of App Use	History of system use, like history of app use and recorded data
	Location Information	Location information measured by GPS devices, location registration information sent to the base station
	Photograph, Video	Photographs and videos shot on smartphones

Source: Ministry of Internal Affairs and Communications of Japan, 08. 2012

2. Theoretical Discussion on Non-PII

As personal information is produced and combined through various channels like smartphones, SNS, and the Internet of Things, personal information is in a gray zone where distinguishing the levels of identity and combination is difficult (See Figure 1). Therefore, the main question is whether "information uploaded by a person" or "personal data" can be considered personal information.

Personal information in Korea is fundamentally based on personal identification. As yet, Korea has no definite standard or agreement on whether personal data such as information that can identify a device used by an individual (e.g., cookie information, IP address), even if it does not identify an individual; information that can identify an individual (e.g., ID) in the service used; information created by an individual (e.g., history of behaviors); and individually created information (e.g., social information) can be included in the category of personal information.

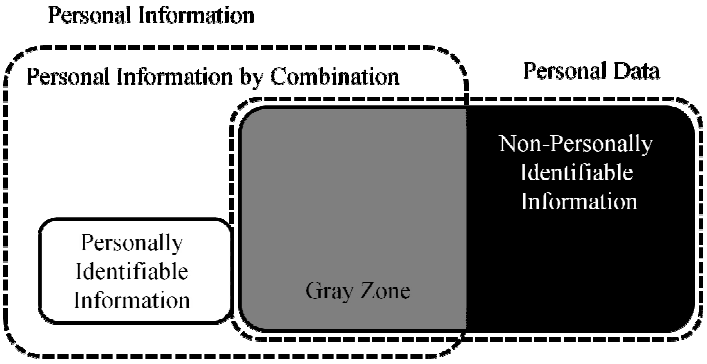

Figure 1: Scope of Personally Related Information

Therefore, we need to discuss and agree on expanding the concept of personal information from being that focused on personal identification to "personally related information," which includes personal data that can estimate the personal state. We must also simultaneously discuss how far personally related information should be regulated if personal information is thus expanded. In other words, the question comes down to how to measure the sensitivity of personal information. For this, discussions on personal identity are examined and standards to classify personal information are located. First of all, identity is the ability to identify the individual concerned by using specific information and is categorized into four types: verinymity, persistent pseudonymity, linkable anonymity, and unlinkable anonymity (Goldberg, 2000).

First, verinymity is information that can practically identify an individual. Government ID (residential ID), social security number, credit card number, and address are included in this type. Email address, IP address, and digital passport number are also defined as verinymity information. Verinymity has two important attributes: linkability and permanence. That is, to reach (linkability) the individual concerned remains mostly without changes (permanence). Second, persistent pseudonymity is information that can estimate the identification of an individual because it is used for a certain period of time. A pen name and a nym server apply here.

Third, linkable anonymity is information that can specify the individual concerned for a moment. Prepaid phone cards and membership cards are included in this type. Finally, unlinkable anonymity is information that can never identify the individual concerned. Information on a purchaser who paid cash and sender information by anonymous remailers are included in this type. On the one hand, the level of determining anonymity is sometimes taken as two elements: reachability that specifies "who the information sender is" and linkability that identifies "if the information sender is the same person" (Information-Technology Promotion Agency of Japan, 2012). Linkability, especially, that connects various data to the individual involved is considered a significant factor. For example, the ID required when initiating a service or registering a membership validates that the subject of diverse activities is that individual. On the other hand, three criteria are suggested in terms of the identification standard for personally related information: identity, linkability, and observability (Peter Hope-Tindall, 2004). First, identity is the criterion that specifies an individual with the information concerned. There can be complete verinymity to complete anonymity (See Figure 2).

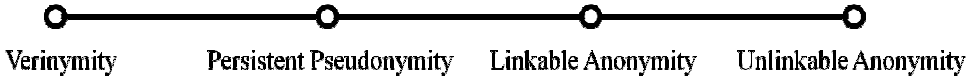

Figure 2: Level of Identification of Personally Related Information

To minimize identity, a design to eliminate specific elements that can identify an individual through system development is needed. Second, link ability is the criterion that measures the relation with the individual through combinations of data elements. For instance, if an individual number is allocated to a telephone card, the phone company can specify the card user's behaviors. For non-link ability, we must remove binding factors, intermediate parameters that provide link ability between pieces of information, that is, removal of the information key. In addition, caution must be observed toward other elements, to proceed with information arranging from time to place and message. Third, observe ability is the criterion that measures level of effect by identity and link ability while using the system. Personally related information is sometimes classified into personal identifier and PII, while PII is categorized into a long term use period and broad use range (Suzuki, 2012).

An example of personally related information with a long term use period and broad use range is UDID; an example of personally related information with a long term use period and narrow use range is a login identifier required by a website. Personally related information with a short term use period and narrow use range is the session ID. In the light of all this, results of examining identification criteria of personally related information show that link ability is the common criterion. Likability determines if the individual can be specified by using information with personal relevance. The next commonly used element, permanence, clarifies individual specificity even more. Permanence means to use information that specifies an individual without changes, and individual-specific error is reduced because this information is used for a longer term. Based on link ability and permanence, the personally related information in Table 1 can be shown in two dimensions (See Figure 3).

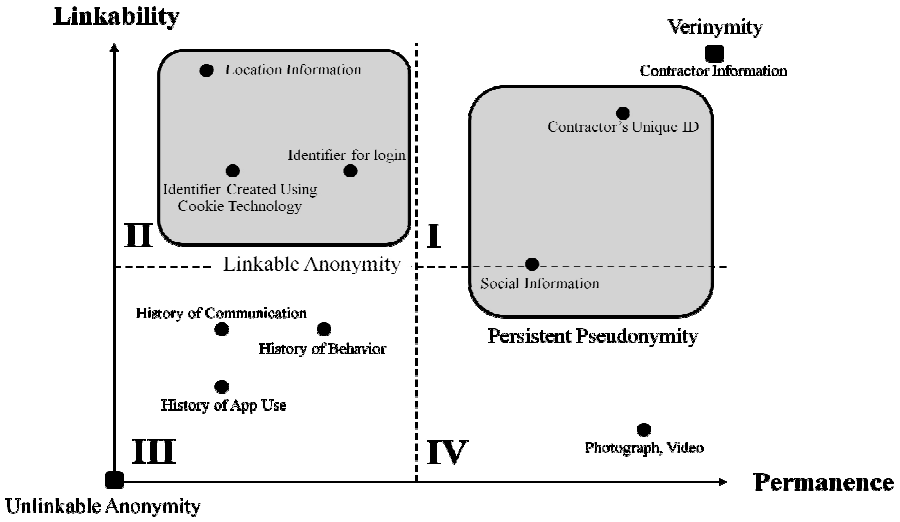

Figure 3: Categorization of Personally Related Information

The contractor's unique ID with high identity due to high link ability and permanence is included in domain I. Social information link ability and permanence are not as high as a contractor's unique ID, but identification by information through personal online social relations and behaviors ranks high. Next, because such identifiers as login identifiers and cookies have high link ability but short permanence they are applied to domain II. Location information, in particular, is positioned on the top left in domain II because it has high link ability but short permanence. Photographs/videos are included in domain IV because they have low link ability but high permanence. Finally, history of communications, history of behaviors, and history of app use have low link ability and short permanence, which places them into domain III. Meanwhile, assuming that the highest point of link ability and permanence is "verinymity" and the lowest point is "no linkable anonymity," personally related information in domain I can be referred to as "persistent pseudonymity" since a brief presumption on the individual concerned is possible. Personally related information in domain II is "linkable anonymity," which can specify the individual for a moment. The information in domain III is close to "unlinkable anonymity."

3. Application of Personally Related Information Types

As stated above, taking systematic protective measures for personal information is possible by categorizing personally related information into verinymity, persistent pseudonymity, linkable anonymity, and unlinkable anonymity, based on link ability and permanence. First, efficient management of personal information is available. With application of these types of personally related information to a personal information system and a relevant database, personal information can be managed safely and efficiently by developing control standards, including information processing by type, fixing utilization level, and preparation of security level. Second, systematizing political regulations on personal information is possible. The current system requires providing protective measures to every piece of personal information identifiable by combination, but the obscurity of non-PII criteria is causing confusion in protective measures. Business operators who deal with personal information as mobile advertisers are strongly required to classify the security level depending on types of personally related information and to take reasonable measures for each level. Third, recognition of the Right to Self-control on Information Management and privacy literacy has improved. As types of personally related information become standards for judging and controlling the degree of an individual's personal information, they help an individual enhance the Right to Self-Control on Information Management and recognition of information protection. It is also possible to raise the effectiveness of the Right to Self-Control on Information Management by suggesting usability of information usage by type, potential risk, and control methods for personally related information, along with companies' personal information protection policies.

References

Aboutmyarea (2011). "Is your Website Cookie-Compliant?" Accessed from http://www.aboutmyarea.co.uk/Central-London/London/SE9/News/Local-News/198826-Is-your-Website-Cookie-Compliant

BBC news (2012). "Government to Miss Cookie Cut-off." Accessed from http://www.bbc.com/news/technology-18090118

FTC (2000a). "Online Profiling: A Report To Congress." " Accessed from https://www.ftc.gov/sites/default/files/documents/reports/online-profiling-federal-trade-commission-report-congress/onlineprofilingreportjune2000.pdf

FTC (2000b). "Online Profiling: Benefits and Concerns." Accessed from https://www.ftc.gov/sites/default/files/documents/public_statements/prepared-statement-federal-trade-commission-online-profiling-benefits-and-concerns/onlineprofile.pdf

Global Regulatory Enforcement Law Blog (2012). "The ICC publishes its 'UK Cookie Guide' on 2 April 2012 to provide guidance to website operators and website users alike."

ICC UK (2012). "ICC UK Cookie Guide." Information-Technology Promotion Agency of Japan (IPA) (2012). "life and economy by IT connectivity."

Ministry of Internal Affairs and Communications of Japan (2012). "Public Announcement of「Innovation in the New Era by Proper Handling of User Information of Smartphone Privacy Initiative and Literacy」."

Peter Hope-Tindall (2004). "Privacy Impact Assessment for e-Government." Ministry of Internal Affairs and Communications of Japan.

Privacyassociation (2011). "New Dutch cookie law requires prior consent from Internet users." Accessed from https:// www.privacyassociation.org/publications/2011_06_28_new_dutch_cookie_law_requires_prior_consent_from_internet_users

Suzuki, M. (2012). "Assignment of Personal Information Protection Act in the Era of My Number System."