ISSN 1883-1974 (Print)
ISSN 1884-0787 (Online)
National Institute of Informatics News

NII Today

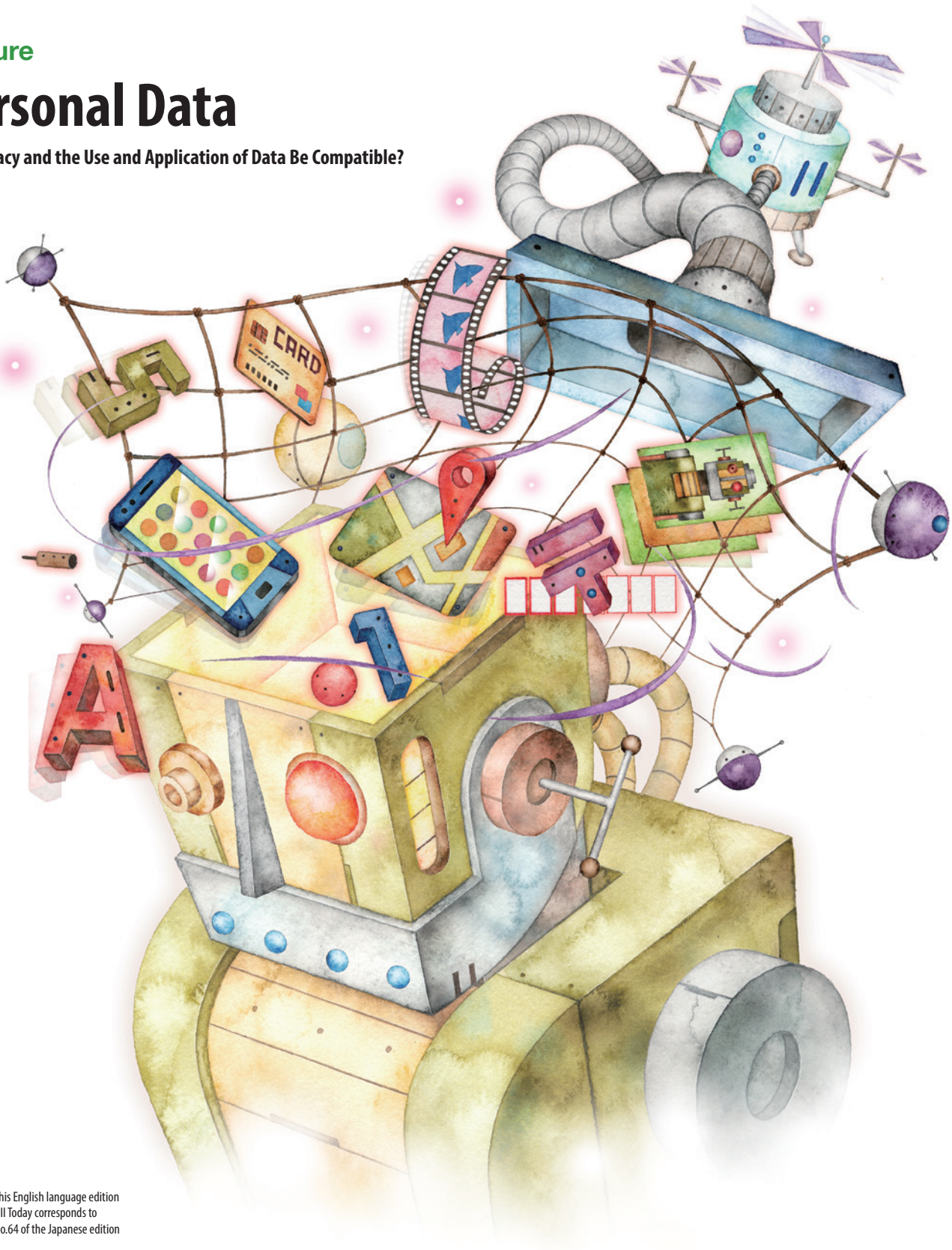
National Institute of Informatics News

50
Jun. 2014

Feature

Personal Data

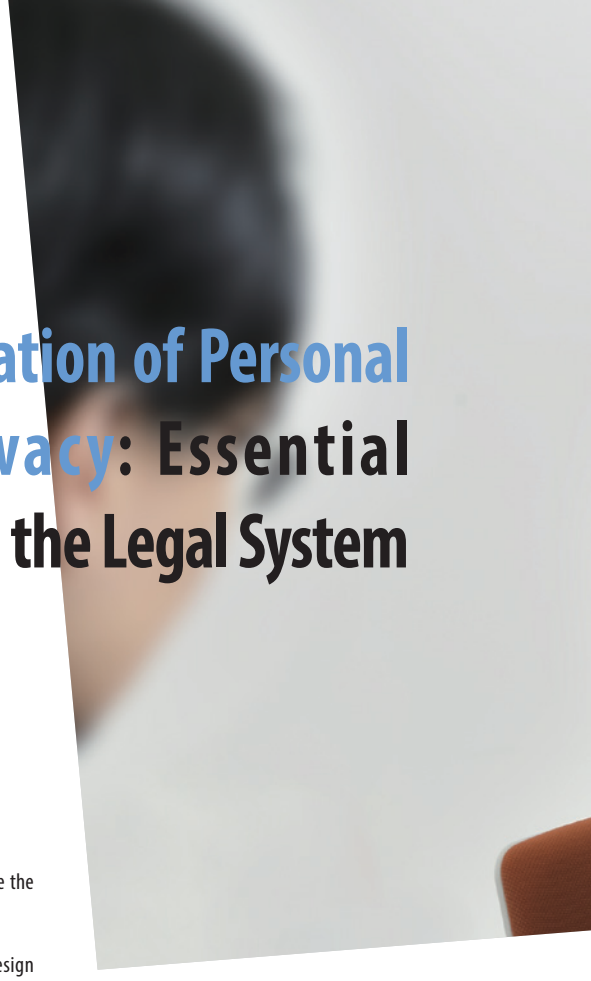
Can Privacy and the Use and Application of Data Be Compatible?



This English language edition NII Today corresponds to No.64 of the Japanese edition

NII Interview

Compatibility of the Use and Application of Personal Data and the Protection of Privacy: Essential Cooperation Between Technology and the Legal System



Ichiro Satoh

Professor, Information Systems Architecture Research Division, NII
Professor, Department of Informatics, School of Multidisciplinary Science, The Graduate University for Advanced Studies

The creation of new value using so-called big data is garnering attention. In particular, personal data that includes information such as personal behavior is highly valuable but also contains an element of difficulty in the protection of privacy. In the Working Group for Technical Issues (hereinafter "WG") of the Study Group on Personal Data (hereinafter "Study Group") of the government's IT Strategic Headquarters, new systems were discussed from a technical standpoint. We interviewed Professor Ichiro Satoh, who serves as the chair of the WG, to hear about the situation surrounding use and application of personal data and his awareness of existent problems.

Hoshi First, would you tell me how you have become the chief administrator of the Study Group and WG?

Satoh The Study Group is a committee that will design a system for handling personal data. It is closely related to technologies including informatics, and whether we protect personal data or use and apply it. So as a specialist in informatics I joined the committee. At the same time I was also involved in the WG, a task force that supports the technical aspects of personal data.

Hoshi Why did you decide to reestablish rules on personal data?

Satoh A basic law for personal data is the Act on the Protection of Personal Information (the existing law) that was established in 2003. Although the review of the Act in 2008 was included in the supplementary resolution, it was not reviewed. There are two further reasons. One is that it is difficult to say that the existing law covers technological progress. Although we already had the Internet and Web search back in those days, the amount of information and search capabilities have greatly increased. The other reason is relationships with foreign countries. Although the handling of personal information was initially strict in Japan, it has become relatively looser as a legal system has developed in European and other countries. If this continues, a problem could arise in that Japanese companies doing business in Europe and the United States will be unable to locate customer databases in Japan. This is obviously detrimental to Japan's national interest.

Hoshi What will become available by the "provision of personal data to a third party" that was discussed especially deeply in the law's revision?

Satoh When using and applying information, a business operator does not always possess necessary information. In many cases, the operator wants to use information of other operators. But this information could contain personal information.

For example, a mail order site recommends products by using the purchasing behavior of other users of the business operator. For users, there is an advantage of being able to quickly find products they want. Making provision of data available to a third party will make it possible to encourage manufacturers to improve their products by informing them of the behavior of the mail order site's users. At the same time, however, from the standpoint of protecting personal information there is a problem in provision of data to a third party.

Hoshi Please tell me about anonymity as discussed in the WG.

Satoh Anonymity is a technology to process information so that it is difficult to identify individuals. When providing a third party with personal data, anonymity is required for appropriate processing of information that enables identification of individuals. The European Union recently established a law that makes anonymized information available. In Japanese law we will also develop provisions for anonymity.

Hoshi Is there any general method for anonymity?

"From the stage of developing a technology, we should start developing an institution considering the future when the technology spreads, in cooperation with legal experts."

Satoh For instance, as to whether it is enough to delete the name and phone number of individuals from data, I would say there is no easy answer. If birth date, sex, or postal code remain in the data there is a 70%-80% probability of being able to identify individuals by comparing with external data, e.g., postcode (zip code), date of birth, and sex. And since the identification depends on the type and characteristics (degree of dispersion) of data and external data to compare with, there is no general method of anonymity and it is impossible to set unified standards. Also, the more strictly we promote anonymity, the more difficult the use and application of data will be.

Hoshi What are the points to keep in mind in anonymity?

Satoh Anonymity requires more sophisticated knowledge and technologies than data analysis. Data analysis and anonymity have a relationship like a hatchet and a shield, and anonymity to protect data is difficult without knowledge about data analytical methods. However, since it will take time to develop human resources for such a job, measures need to be developed assuming a limited number of such human resources. The model of providing data to a third party in the WG is based on this assumption.

Hoshi What challenges will arise if the use and application of data are even more active in the future?

Satoh I am particularly worried about facial recognition and genomes (genes) because advanced technologies have dramatically improved identification speed and accuracy for both with low costs. For example, privacy will become nearly visible in some cases if headshots from security cameras on a street corner and social network images are compared, no matter how anonymized other data are. Nevertheless, if we overregulate security cameras, safety could be endangered.

Hoshi How should we protect privacy in the years to come?

Satoh These days, technology called big data is becoming a major theme. In big data, combining data for differences will draw out new knowledge. On the other hand, comparison of data will cause a new privacy problem. For example, by comparing consumption of water and gas will make it clear how long a person takes a bath.

Hoshi How should we protect personal data?

Satoh As technical experts we have no choice but to acknowledge this reality. It is not limited to personal data. There are also an increasing number of cases in which problems caused by technology cannot be solved by technology alone. In fact, new privacy problems are created partly by the advance of technologies such as big data and facial identification. It is not always true that there is a technology to fundamentally prevent it. If this is the case, we have to solve problems by means other than technology, such as with an institution applying disciplinary rules. In the provision of personal data to a third party as proposed in the WG, we try to protect personal data by combining technologies with institutions, such as imposing disciplinary rules for sources and receivers of personal data not to identify individuals from data provided, instead of requiring complete anonymity of the data provided.

Hoshi How do you think the relationship between technologies and institutions will play out in the future?

Satoh In the future we should think that technologies and institutions are inseparable. In other words, from the stage of developing a technology, we should start developing an institution considering the future when the technology spreads, in cooperation with legal experts. Conversely, since an institution itself relies on technologies in an increasing number of cases, opportunities for technical experts to participate in institutional design will increase. For example, wearable cameras have been in the news lately, and we should propose a legal system to minimize problems together, rather than for simply selling products, in expectation of social problems associated with the use of the products. Tax collection and criminal investigation also rely on IT. I believe that informatics researchers will be required to make an institutional design and social design suited for technologies in anticipation of future technological development. In that sense, the role of NII, which is a research institution dealing with informatics, should grow.

A Word from the Interviewer

In the new system under discussion, anonymity will be obligatory in providing a third party with personal data and be placed under the surveillance of a third-party institution. Although some say that this will increase the burden of the business community, Professor Satoh points out that we should not be left behind global trends. There is a tradeoff (reciprocity) relation between the advantage of the use and application of personal data and the protection of privacy. Information technology has deeply permeated every part of society. I sensed that the knowledge of technologies and institutions has become increasingly necessary for general citizens and consumers.

Akio Hoshi
IT journalist

Akio Hoshi graduated from the Graduate School of Science and Engineering, Waseda University. He went independent in 2006 after working as a writer for Nikkei Electronics and editor in chief of the Nikkei Java Review, an online magazine, at Nikkei BP. He has a broad variety of coverage experience ranging from semiconductors to operating systems, programming languages and Internet services.

Toward Revision of the Act on the Protection of Personal Information

Clarification of Distinguishability and Establishment of an Independent Third-party Institution Are the Keys

That's **Collaboration** 1

Currently, efforts are progressing to revise the Act on the Protection of Personal Information, which has been fully enforced since 2005. Behind these efforts is an aim to use and apply personal data while protecting privacy, in addition to revising the law in a more practical form along with the rapid advancement of technology. To learn about problems in the existing law and expectations and challenges for its revision, we interviewed Professor Hisamichi Okamura, a lawyer known as an authority on the Act and the leading expert on legal issues regarding the Internet and information security.

Problems such as overreaction, technological advancement and the competent minister system

Satoh Could you tell me what problems we have in the existing law?

Okamura Overreaction became a problem at the beginning of enforcement of the law. For example, in the derailment accident on the JR Fukuchiyama Line, since some hospitals to which victims were taken refused to answer the safety confirmation from their families, a great deal of confusion was created. I fear that some families were unable to be with the victims when they passed away. There were also many bereaved relatives who wanted to know the contents of victims' cell phones, but this wish was blocked by "secrecy of communication" guaranteed by the constitution. On the other hand, incidents in which information that should be protected is leaked show no sign of significant decline. I am working on this issue with a desire to correct this disjointed situation and achieve a balance between protection and distribution of personal information.

Satoh Under such circumstances, you are involved in the policy planning of central government offices and speak at public lectures and appear in the media as a visiting professor at NII.

Okamura Yes, I do this because there are only a few academics involved in the legal system in the information field. Moreover, there are few attorneys who can take a bird's eye view in speaking on the law from its drafting to the current situation. I think that it is also my role to explain these new problems and relationships with the legal system to the general public in an easy-to-understand manner.

Satoh Other than the overreaction you mentioned, what problems are there in the existing law?

Okamura Broadly, there are three types of problems. The first is that the social system does not catch up with technological advancement. The second is that guidelines are vertically segmented in each area since this act adopts the competent minister system[*1]. Even at an international meeting[*2] of the OECD it is difficult to make remarks in the national interest, since Japan sends not a commissioner but an observer to the meeting. New technologies that cannot be dealt with by a vertically segmented system have also come into existence. For example, the guidelines of the Ministry of Internal Affairs and Communications

Hisamichi Okamura
Visiting Professor, NII
Lawyer

are applied to cell phone companies, but for Osaifu Keitai, an e-money function, and Mobile Suica, an IC travel card service, both on cell phones, the guidelines of the Financial Services Agency and the Ministry of Land, Infrastructure, Transport and Tourism are respectively applied. As such, all guidelines are applied redundantly. Ultimately it is necessary to create a third-party institution as the core supervisory organization. The third problem is that the concept of "individual distinguishability" is extremely difficult to understand. I think that if we clarify this concept, we will also be able to deal with the problem of overreaction I mentioned.

Redefining the vague concept of "distinguishability"

Satoh Distinguishability also became a topic of focus at the Technology WG of the Study Group on Personal Data, where I served as the chief administrator. Given that distinguishability changes according to technological advancement, is it necessary to review the concept on a case-by-case basis?

Okamura You make an important point. However, since distinguishability follows the concepts in the OECD Guidelines and the EU Directive, it is necessary to clarify the concept definition from an international perspective in the modern world in which data are exchanged internationally. And given that the vast amounts of ordinances, laws and regulations, including grounds for non-disclosure in the information disclosure system[*3] and the Act on Use of Numbers[*4], follow the concept of the Act on the Protection of Personal Information, it is necessary to work on this issue carefully, taking consistency into account.

Also, as indicated by the fact that even if a defendant is declared guilty at the first trial he is found innocent at the court of second instance, the law contains the aspect that the result changes depending on the way of looking at the case. Or more than one low may apply a case, and we may see a situation in which even though a case does not violate the Act on the Protection of Personal Information it infringes on the right to privacy. I think that we need to go into greater depth in the revision of the law to consider what actual harm will arise if the "right to control personal information" is infringed upon, as well as to develop the law so that the Act and the right to privacy will not be contradictory.

Expectations for a Third-Party Institution and Development of Human Resources

Satoh You are considering establishment of a third-party institution associated with the revision of the law. What are its advantages and disadvantages?

Okamura An advantage is that we will be able to respond to cross-sectoral problems. In the current conditions, 40 or more types of guidelines stand together in government agencies, but they are no longer capable of dealing with actual business in the vertically segmented system. It is also an advantage that we will be able to argue our position to foreign countries as the "privacy commissioner" of Japan, in light of the overseas transfer of data. I think that the extended and developed form of the Specific Personal Information Protection Commission[*5] currently established in the Cabinet Office would be an appropriate image of the institution. From the standpoint of international harmonization there are many problems to solve, including whether a supervisory organization over public sector agencies should be a highly public institution, how much authority it should be delegated, what penalties (surcharges) should be imposed if it violates rules, and how to secure human resources.

In any case, Japan will revise its domestic law for the first time in light of the revised OECD Guidelines, I want those involved in the revision to work on it with the spirit of taking the lead internationally when finalizing the broad outline. However, as technology advances rapidly we should bring a review in the future into view, instead of making a one-shot decision. We should have great hopes for the birth of the Japanese version of the privacy commissioner.

Satoh I feel vexed that we cannot solve technical problems through technology, but don't we have to solve them in close cooperation with institutions going forward, with technical experts and attorneys working hand in hand?

Okamura As technologies to maintain the status quo and those to break it are always developed, playing a cat-and-mouse game, it is essential to follow up technologies from an institutional standpoint. However, compared with the speed of technological advancement, the law, by nature, falls behind the curve no matter how hard it tries. From that perspective I think that attorneys and technical experts have to draft measures jointly.

On the other hand, there are situations where some attorneys don't know about matters in different fields if their specialty differs. At the basis of the protection of personal information lies the right to privacy that has been admitted by judicial precedents, and this right forms a pair with "freedom of expression" in the constitution. I think we should essentially aim to achieve a good balance between freedom of information distribution in a public space and control of self-information in a closed private space, yet some attorneys have knowledge of personal information but are out of touch with other legal systems. Therefore there is an urgent need to develop human resources capable of designing institutional arrangements while staying alert across the board. In that context, NII needs to cultivate both technical experts who have knowledge of law, and attorneys who have knowledge of technology, and develop as a valuable space where both can have beneficial discussions.

(Interviewed by Ichiro Satoh, Written by Madoka Tainaka)

*1 Competent minister system – Each competent minister supervises individual areas under his/her jurisdiction by providing guidelines, etc. according to the actual condition of the areas.

*2 International meeting of the OECD – At the Council of the Organisation for Economic Co-operation and Development (OECD) in 1980, the Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data was adopted (revised in 2003). The eight principles stated in these guidelines have become standards for the legal system to protect personal information of OECD member countries.

*3 Grounds for non-disclosure in the information disclosure system – Information on individuals is stipulated as non-disclosure information in information disclosure laws and ordinances.

*4 Act on Use of Numbers – Abbreviation for the Act on Use, etc. in Administrative Procedures of Numbers to Identify Specific Individuals

*5 Specific Personal Information Protection Commission – A third-party institution of an extra-ministerial bureau of the Cabinet Office, of which the mission is to take necessary measures to ensure appropriate handling of individual numbers and other specific personal information, while paying attention to their availability

NII Special

1

For Flexible Use and Protection of Personal Information

Revealing Attitudes of People Through Conjoint Analysis

People's awareness of the use and protection of personal information varies. Although the basic idea that personal information should be protected is commonly shared, the opinion about how much information should be provided varies considerably from person to person, and attitudes change significantly depending on what can be gained in compensation. Revealing such attitudes using conjoint analysis is one of the research themes focused on by Associate Professor Hitoshi Okada. The research results are being applied to rulemaking for protection of personal information.

A flexible approach is necessary to protect and use personal information

Discussions grow active on the use and protection of personal information.

Divulging of personal information from companies and cybercrimes such as defrauding of corporate information and e-money using leaked personal information are taking place. As is widely known, issues related to personal information have been a global concern along with the diffusion of IT.

However, Okada warns that it is meaningless to be preoccupied simply with the discussion of whether or not personal information should be protected.

People have different opinions about personal information.

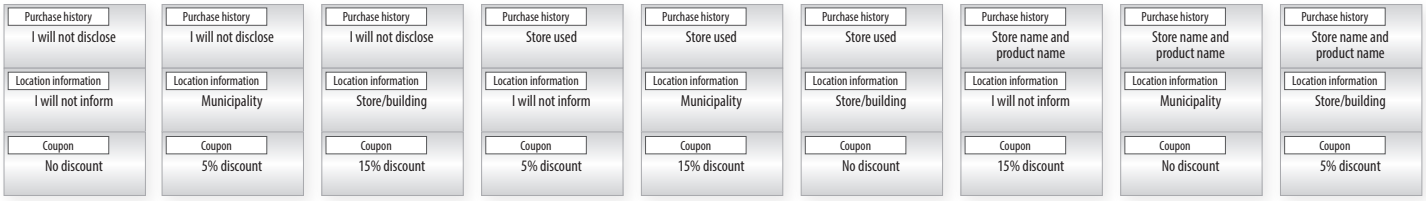
Some say that they do not want to provide their personal information at all, and others say that they do not mind providing their name and age but do not want to provide any other information, or they do not mind disclosing information up to their purchase history or location information. Their opinions also vary widely depending on services to be provided.

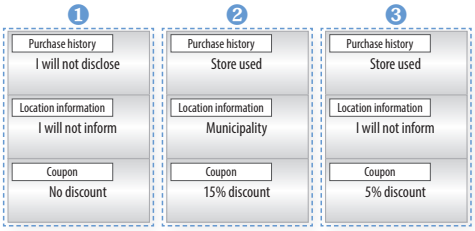
There are a considerable number of cases where people unintentionally provide personal information when they are accessing services. For example, a user of social media who does not want to provide personal information uses the social media to widely disclose the atmosphere of the place he is eating. We could say that this is a case where the awareness of personal information and actual behavior differ.

On the other hand, only giving priority to the protection of personal information institutionally could become a factor that hampers creation of new services using information because restrictions on the use of personal information will increase. For companies, obtaining unnecessary personal information will only be a risk, but it is obvious that collection and use of appropriate information will help create innovative services and increase their business competitiveness. Forward-looking discussions are necessary to promote the use of personal information, while achieving a balance with protection.

"When we discuss use and protection and the rulemaking of personal information," Okada says, "it's dangerous to have discussions only from an alternative perspective. A more flexible perception is needed."

In an example of the screen of the conjoint research survey, nine cards are visible on the Internet screen.
When a respondent clicks them in the order of preference, the order is recorded.





❶ Card that appears to be selected as first by a person who does not want to disclose anything

❷ Card that appears to be selected as first by a person who does not mind actively informing if there is some benefit

❸ Card that appears to be selected as first by a person who does not mind providing his/her purchase history but is unwilling to provide his/her location information

Hitoshi Okada

Associate Professor, Information and Society Research Division, NII
Associate Professor, Department of Informatics, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies

Conjoint method to elicit users' real intent

Okada uses conjoint analysis to study the individuals' awareness of protection and use of personal information as described above.

Conjoint analysis is an analytical method to reveal people's ideas and behavior by acquiring correlations with elements that are in a tradeoff relationship by showing multiple evaluation items to respondents and repeatedly asking them about these items.

In the study Okada is working on, randomly selected subjects compare nine kinds of cards displayed on the screen, using the Internet, and select those that match their attitude toward personal information in sequence. He began the study in 2010 and performs analysis from various angles while changing the items to set. At the most frequent rate he collects information once a month.

In a general survey, if we ask, "Do you think that personal information should be protected?" nearly all respondents will answer "Yes." And the response "I don't want to provide personal information more than necessary" should also be frequent.

However, if items such as "Your personal information will be used for governmental support at the time of a disaster" and "Your personal information will be destroyed after a certain period" are added to "I will provide my personal information," it is believed that many respondents will agree to provide it. In other words, conjoint analysis clarifies the attitude of people by showing the subjects cards that combine various elements and repeatedly researching which card is prioritized.

"Even those who think that they don't want to provide personal information will answer 'I will provide some personal information' if they can gain a benefit such as reward points that can be used for shopping by providing personal information. People also post pictures or comments, which can be considered personal information, on social media because they feel a benefit of enjoying the service. In my study there were also many responses showing unwillingness to provide personal information when using a travel card, while many responses showed willingness to provide personal information when using a card issued by a retailer. This suggests that card users look for some sort of benefit. The attitude of consumers toward provision of personal information changes significantly depending on the situation, use and what benefits they can get."

Okada describes this as the "tradeoff between costs and benefits."

"Consumers are keenly aware of how much service they can get for the information they provide. A mechanism of regarding personal information as a cost and receiving services worth the cost as a benefit is starting to be built."

Clarifying costs and benefits and utilizing for rulemaking

If both the providers of personal information and the receivers who use it understand the tradeoff relationship between costs and benefits, it will make a shortcut to construction of the social environment where personal information is used efficiently.

Okada points out that an alternative perspective toward the protection of personal information is not appropriate, saying, "What is the rough standard about how much personal information individuals will provide in exchange for how much benefit they can receive? And how much service do companies provide for how much information provided? The mechanism that a contract is made between the parties based on these mutual understandings should be constructed."

Currently, the government plans to outline the draft revision of the Act on the Protection of Personal Information by June 2014 and submit the draft revision to the regular Diet session in early 2015 after public comment.

The draft revision will include some new rules on the handling of personal information (personal data), and these rules will evolve into more flexible ones that presuppose that a contract between the parties will be the basis for providing personal information, among other things.

Originally, conjoint analysis has often been adopted when new rules such as laws and ordinances are made. The result of Okada's study using conjoint analysis for personal information is likely to contribute to realization of a non-alternative type of rulemaking for protection of personal information.

(Written by Katsuyuki Ohkawara)

NII Special

2

Get Control of Personal Data Back in Our Hands

PrivacyVisor raises discussion on arbitrary facial recognition

The risk that personal data is used for unexpected purposes due to collection of big data and development of analytical technologies has started to be pointed out. The situation in which personal information, such as the name and address of a person in a picture, is entirely exposed independently of the person's intent from the content registered or posted on an SNS has become a reality. Is it possible to control personal data in cyberspace by one's hands? The PrivacyVisor, a tool for preventing facial detection, developed by Professor Isao Echizen, is causing a new stir in discussions on the use of big data and the protection of privacy.

Problems in Facial Recognition

Associated with the popularization of cell phones with cameras and smartphones and the expanding use of SNS, snapshots on the street and of commercial facilities are increasingly disclosed on the Internet. Even if permission is obtained from a person photographed, those who happen to appear in the picture behind the person may not have thought their pictures would be made public. Using current facial recognition technology, however, the same people can easily be identified from facial features of those in the picture. And even the date and location information can be disclosed through the information incidental to the picture. If a person can be identified, it will not be difficult to search and collect related information from information sources on the Internet. In other words, someone unknown to you will come to know what you do, when and where from a single picture in which you happened to be photographed, and even other information such as your job and friends will be revealed if various links are followed.

In addition, since practical application of wearable devices with cameras, as represented by Google Glass, has been imminent, times when the personal data of a person viewed through these devices can be revealed in real time and onsite are just around the corner.

Professor Isao Echizen at NII says, "I want people to think about what they can do to protect their privacy against the fact that their own information will spread in the cyberspace and could be used for a commercial purpose. For example, when Carnegie Mellon University experimented on whether those who agreed to have their picture taken anonymously could be identified based on their headshot by checking with Facebook, one-third of the test subjects were identified, and even the personal interests and social security number of some test subjects were revealed. Times when privacy is exposed only by a headshot have already arrived."*

Isao Echizen
Professor, Digital Content and Media Sciences Research Division, NII
Professor, Department of Informatics, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies

PrivacyVisor tricks facial detection

Given the concern that infringement on privacy could become a reality precisely because we are living in the era when big data are available, Echizen continues rapid development of the PrivacyVisor. "The PrivacyVisor," he says, "is an attempt to protect ourselves by preventing our facial image, which is personal information, from being used arbitrarily. With the device, we will be able to prevent infringement on our privacy when photographed by nullifying facial detection, which is a preprocessing of facial recognition, whatever the mechanism of those who collect information may be."

Currently, even standard digital cameras carry a technology for detecting only human faces from images. The Viola-Jones method, adopted as a typical facial detection method, detects a face by checking with the accumulated and learned image data of a face by extracting differences in luminance (light and dark) mainly around the eyes, bridge of the nose and the nose itself.

The method is fairly accurate and it is said that not only faces photographed from 22 meters away, but also in an experiment by Echizen even those disguised with five types of sunglasses of different design could be detected accurately. Although measures to force this facial detection to fail have been proposed, such as coloring the face with special pigments, using a special hairstyle, or wearing a mask covering the face, there was a weak point in the measures: that is, they pose an obstacle to face-to-face communication in the real world. As a tool to prevent facial detection that can be used more naturally, the goggle-type PrivacyVisor was developed.

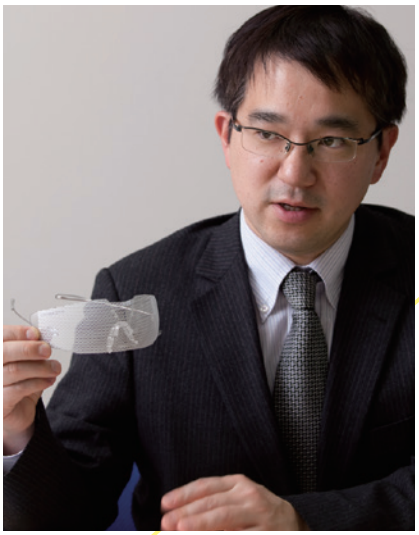
Echizen installs 11 near-infrared LED lights in goggles that cover areas around the eyes so that users can put the lights on when they do not want to have their face detected. Since near-infrared lights have a wavelength beyond the human visible range, they are invisible for a person facing the user. However, because consumer-use cameras detect the infrared range, LED lights are reflected as noise. As a result, luminance differences around the eyes cannot be detected, so facial recognition will fail. This is how the PrivacyVisor works.

In an experiment shooting 10 test subjects standing in front of a camera, the facial detection of seven to eight of them who stood 20 meters away succeeded when they did not wear the PrivacyVisor. While the facial detection rate remained almost unchanged when they wore the PrivacyVisor but did not turn the LED lights on, facial detection was not successful for any subjects standing within 22 meters if the lights were on.

New and fashionable PrivacyVisor needs no power source

The trial product had trouble in that a battery was needed to light the LEDs. And it failed to work since professional cameras did not sense the infrared range. Therefore, since the end of last year and jointly with eyeglass manufacturers in Sabae City, Fukui, which is recognized worldwide for its eyeglasses, Echizen has been moving forward with development of a new PrivacyVisor that needs no LED lights.

Prototype of the PrivacyVisor



The new visor uses a material that reflects visible light for concave areas around eyes, etc. that a face detector deems dark, and conversely a material that absorbs visible light for convex areas deemed bright, such as the upper part of the nose. With white as the basic tone, a trial product sprinkles fine dots and patterns in its transparent parts. This is likely to relatively reduce the sense of discomfort and burden when carrying the device and wearing it like sunglasses. Echizen and the eyeglass manufacturers are said to be seeking launch of a prototype by year's end after adding some design improvements.

Balance between use and application of information and protection of privacy is important

Meanwhile, one concern is the possibility of the PrivacyVisor being used for criminal purposes.

"For example, facial recognition is used in security cameras at airports for identifying terrorists," Echizen says. "In such places, institutional measures such as not allowing the visor to be worn will be necessary. It is important to achieve a good balance between public interest and protection of privacy." The PrivacyVisor also presents this challenge.

Echizen also points out the possibility that wearing the PrivacyVisor could become a sign of refusing use of personal images. He proposes a mechanism to promote the use and application of data by indicating the permissible scope via what we wear (such as accessories and badges) and based on the conditions thereof.

In the age of big data, it seems necessary to think carefully about what to protect and disclose in our personal data to minimize risk and maximize our interest, as well as how to protect data.

(Written by Masahiro Doi)

*Following a request from the European Union (EU), Facebook decided in 2012 not to provide the facial recognition function in Europe. Google Glass has also expressed that it will not be equipped with the function. However, as long as such technology remains practicable, new service providers or application developers could develop and provide facial recognition functionality, even if Facebook and Google do not.

Balancing Utility with Anonymity of Data

Latest Trends and Challenges in Anonymization Technology

That's **Collaboration** 2

Anonymization technology is essential for achieving protection of privacy when using personal data. How deeply has anonymization technology evolved? And what challenges have emerged for actual operation? Professor Ichiro Satoh at NII (the chief administrator of the Technology Review Working Group [hereinafter "Technology WG"] of the Study Group on Personal Data of the government's IT Strategic Headquarters) interviewed Katsumi Takahashi of NTT Secure Platform Laboratories, a member of the Technology WG, to hear about the latest trends in anonymization technology and the outlook and challenges for its practical use.

What is anonymization in the big data age?

Satoh First, could you tell me about "anonymization," which is the key for use and application of big data in the future?

Takahashi The basic definition of anonymization is the process of preventing individuals from being identified by deleting and/or changing information such as names, birth dates or addresses included in personal data. In other words, we call technologies and methods to increase the anonymity of data and the combination thereof anonymization. However, since we are in the age of big data and a great deal of information has been accumulated in the world, there are issues wherein individuals are identified by matching with other data even if only the name and address are deleted.

Anonymization in big data is requested under the current interpretation of the legal system, behind which are a variety of needs for free use of personal data by making them anonymous. In many cases, however, it is not easy to process personal data into non-personal data. Some may say that all we have to do is use anonymization technology, but they lack understanding of this technology. Naive anonymization such as deleting names in data is relatively easy and in the world this is what the term commonly refers to, but it is impossible to eliminate the risk of individuals being identified only with naive anonymization.

Satoh So you are saying that anonymization cannot be accomplished simply and uniformly?

Takahashi That's right. On the other hand, it is possible to process personal data into a state that has almost entirely eliminated individuality included in data. Statistical data is a good example. In this case, however, since the level of abstraction of data is too high, the data will often become unsuited for analysis. This is a dilemma between the anonymity and utility of personal data.

Technology that ensures privacy without destroying the utility of data

Satoh What should we aim for to overcome the dilemma?

Takahashi A theoretical goal is to leave only the necessary minimum amount of information according to the analytical purpose. In the Technology WG, we expressed as a basic concept of anonymization: "There is no such thing as versatile anonymization method. It is on a case-by-case basis according to the types, features and utilization purposes."

Under such circumstances, it will not necessarily be a fundamental issue whether or not information is personal. Rather, it will be important to adopt the best data processing method according to the nature of personal data and the purpose of the analysis. It also needs to be ensured that data will be processed safely, including the provision of institutional security on the user side by accurately expressing risk regarding privacy to the information providers.

Ichiro Satoh
Professor, Information Systems Architecture Science Research Division, NII
Professor, Department of Informatics, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies

Katsumi Takahashi
Executive Research Scientist
General Manager of Information Security Project
NTT Secure Platform Laboratories,
Nippon Telegraph and Telephone Corporation

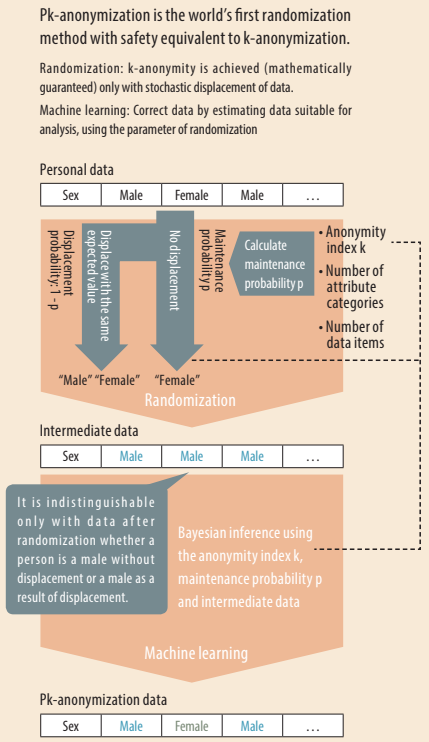
Figure 1. Example of k-Anonymization

Membership number	Birth date	Address	Age	Item purchased
1001	April 1, 1979	A-cho, Chuo-ku, Tokyo	34	Bread
1002	December 10, 1986	A-cho, Yokohama-shi, Kanagawa-ken	26	Comic book
1003	October 10, 1974	B-cho, Shibuya-ku, Tokyo	38	Ice
1004	May 5, 1991	B-cho, Kamakura-shi, Kanagawa-ken	22	Paperback book
1005	November 10, 2006	A-cho, Kawagoe-shi, Saitama-ken	17	Cola
1006	February 6, 1990	C-cho, Atsugi-shi, Kanagawa-ken	23	Timetable
1007	August 15, 2003	B-cho, Urawa-shi, Saitama-ken	19	Milk
1008	September 30, 2000	C-cho, Omiya-shi, Saitama-ken	9	Tea
1009	January 1, 1983	C-cho, Nerima-ku, Tokyo	30	Boxed lunch
1010	July 7, 1994	D-cho, Yono-shi, Saitama-ken	18	Water

State Fulfilling K-Anonymity (k=3)

Membership number	Birth date	Address	Age	Item purchased	
1001	April 1, 1979	Tokyo	30s	Food	3
1003	October 10, 1974	Tokyo	30s	Food	
1009	January 1, 1983	Tokyo	30s	Food	
1002	December 10, 1986	Kanagawa	20s	Book	3
1004	Deletion 5, 1991	Kanagawa	20s	Book	
1006	February 6, 1990	Kanagawa	20s	Book	
1005	November 10, 2006	Saitama	Minor	Beverage	4
1007	August 15, 2003	Saitama	Minor	Beverage	
1008	September 30, 2000	Saitama	Minor	Beverage	
1010	July 7, 1994	Saitama	Minor	Beverage	

Figure 2. Conceptual Diagram of Pk-Anonymization



Satoh What methods do we have as technology?

Takahashi An example is that if we want to handle personal data as is for an analytical purpose, ciphers are useful. We have been studying and developing secret sharing and secure computation technologies to use for processing data while they are encrypted in order to achieve protection of privacy and safety management measures. This group of technologies is a tool to pursue the confidentiality of data to the utmost limit without impairing the data's accuracy.

On the other hand, if we place emphasis on anonymization, we also have a method called "k-anonymity," which is a parameter to express the risk regarding privacy. This is an indicator for evaluating data's anonymity, and means the state in which there are always k or more people with a similar attribute. For example, if there are at least 10 people included in the target attribute, whether they are in their 20s or 30s, the anonymity of the data is expressed as "k = 10." In other words, the larger the k value, the smaller the privacy risk. A technology that processes personal data to achieve k-anonymity is k-anonymization, which can be achieved through generalization that makes the value of an attribute rougher or there is deletion of data for a low number of attributes, so that there will be at least k records with the same attribute combination (Figure 1).

Satoh You are making more advanced efforts at NTT Secure Platform Laboratories.

Takahashi Our team has developed a method called Pk-anonymization. This makes data incomprehensible in terms of who they belong to through randomization[*1], which is processing to change individual data probabilistically. In randomization in this method, records are processed to be identified with a probability of 1/k or less. We call this nature Pk-anonymity (probabilistic k-anonymity). After that, we execute processing to estimate the original state of the data by using a machine learning method called Bayesian inference[*2] (Figure 2). By doing so, practical anonymous data for analysis will be constructed. We might say that this is pseudo-personal data based on actual personal data. We think that Pk-anonymization is effective for anonymization of big data while retaining a similar nature to k-anonymization.

Satoh Can we use Pk-anonymization for any data?

Takahashi It is particularly effective for anonymization of long personal data with a number of items. For long data, I believe that order-made anonymization, which conducts anonymization by selecting necessary items from data, is repeated in many cases, but the possibility of k-anonymity being damaged by checking with multiple anonymized data has been pointed out. To the contrary, since Pk-anonymization is resistant to the loss of k-anonymity, it is possible to conduct order-made anonymization repeatedly and process personal data with many items into highly valuable data for analysis while protecting privacy. These are the features of Pk-anonymization.

Protecting privacy both technically and institutionally

Satoh I think that k-anonymity will be greatly involved in a variety of scenarios in the future, associated with the revision of the Act on the Protection of Personal Information, etc. Do you think that k-anonymity will be one of the indicators for deciding to license the provision of data to a third party? And in that case, is it possible to say that provision of data is safe if the k-value is a certain number?

Takahashi That's a difficult question. There is still an issue of whether safety can really be judged only with the k value, and I think that there are some cases in which there is little risk even if k is one. I think that the adequacy of the k value will be determined in its social operation, including the development of the legal system in the future.

Satoh So you are saying that for the appropriate use and application of anonymized data in the future, it is necessary to promote it, with the technology and the institution working as a pair of wheels?

Takahashi For personal data, I think it will be enough to deal with it obeying the legal system. Also, we will be able to freely use data that are clearly separated from individuals. The future challenge is how to deal with "anonymized data" other than these two types. Protection of data privacy will be possible by establishing social rules for the use and provision of anonymized data. Data with a risk regarding privacy can be provided only to a person who can be trusted. In other words, I imagine that problems concerning anonymized data will be gradually solved by establishing and operating the legal system, such as permitting provision of data to a person who observes the rules, but declining permission or imposing punishment if the rules are not observed.

(Written by Hideki Itoh)

*1 Randomization: An anonymization method to add noise data to the original data to the extent to which the addition will not have an impact on data analysis

*2 Bayesian inference: A stochastic method to estimate the probability of an event that causes an observed event from the observed event

NII essay

Lessons from the Past

Hisamichi Okamura
Visiting Professor, NII
Lawyer

Everyone has days they cannot forget. For me, April 25, 2005 was such a day.

On that morning I was on a commuter train bound for Osaka. With the beautiful spring sun shining it was a routine morning not particularly unique in any way. However, immediately after that, the news was filled with stories about something grave. An unprecedented train accident in which a rapid train derailed before JR Amagasaki Station, killing 107 people, had occurred. Ironically, this was the rapid train running right after the commuter train on which I had been riding.

That afternoon, when I entered the room of the legal department chair to have a meeting after I had finished a lecture at Kinki University Law School, the department chair was urgently talking on his cell phone. He received a call from a parent of a law student, asking him to make sure the student attended the class since the parent was unable to get in touch with the student, who could have been on the derailed train. Although the department chair had related people at the university check, they could not confirm that the student attended the class. Many times they called the student's cell phone number, which the parent had given them, but they were unable to get through. When I left the building, a cold drizzle had started to fall. It was later found out that the student had had the worst possible outcome.

Reports at that time said that at the accident site cell phones of victims were ringing repeatedly with calls and emails from family worried about their loved ones' safety. The other side of the phone had actually divided light and dark.

Meanwhile, the injured, reaching several hundreds, were taken to hospitals in a wide area from Kobe to Osaka. Although their families, worried about the victims' safety, visited hospitals where the victims were taken, there were frequent cases in which they were not even told whether the victims were there. This was because the hospitals, afraid of violating the Act on the Protection of Personal Information (hereinafter the "Protection Act"), which had been enforced just 25 days earlier, wavered on disclosing the names of their inpatients. As a result, there appeared to be some families unable to exchange final words with their loved ones. In the wake of this incident, overreaction to the Protection Act has become a social issue.

Nine years after this accident, the study for the revision of the Protection Act is now being advanced. This is to respond to the new use and application of personal data associated with the advancement of information and communication technologies. However, drastic measures against the overreaction still have yet to be taken.

While the future continues in fan-like fashion, with the past as its starting point, events in the past always teach us valuable lessons for the future. Looking back on these occasionally, we must always keep asking about the ideal situation of personal data, thinking how the protection and the use and application of personal data should function.

Weaving Information into Knowledge
NII

National Institute of Informatics News (NII Today) No. 50 Jun. 2014 [This English language edition NII Today corresponds to No. 64 of the Japanese edition]
Published by National Institute of Informatics, Research Organization of Information and Systems http://www.nii.ac.jp/
Address: National Center of Sciences 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
Publisher: Masaru Kitsuregawa Cover illustration: Toshiya Shirotani Photography: Seiya Kawamoto, Satoshi Akashi
Copy Editor: Madoka Tainaka Production: Nobudget Inc.
Contact: Publicity Team, Planning Division, General Affairs Department Tel.: +81-3-4212-2164 Fax: +81-3-4212-2150 E-mail: kouhou@nii.ac.jp