NRI 未来創発
Dream up the future.

野村総合研究所

# FY2016 Study on Combining Appropriate Protection of Privacy with Social Use of Location Data

# Report Summary

March 2017

NOMURA RESEARCH INSTITUTE, LTD.

# Contents


1. Study outline
2. Study findings
   - ① Use case-based investigation of “sufficient anonymization” and methods of producing anonymously processed information
   - ② Investigation of standards for security control measures for information on processing methods, etc.
   - ③ Relationships between “sufficient anonymization” and methods of producing anonymously processed information
   - ④ Survey of acceptability to users
   - ⑤ Compilation of rules on location data handled by telecommunications carriers
3. Future directions

# 1. Study outline

1. Background and objectives
   - The “Location Information Privacy Report” released in FY2014 stated that empirical research and testing should be conducted to investigate standards of “sufficient anonymization” and to evaluate and verify the suitability of methods of processing and management and operational arrangements for the handling of location data protected by the principle of secrecy of communications. A study looking into specific use cases was therefore conducted in FY2015.
   - Building on past findings, the present study considered a wider range of use cases in order to conduct a more in-depth analysis of “sufficient anonymization” and investigate anonymously processed information as provided for by the amended Personal Information Protection Act. Rules were then compiled based on the findings.

2. Main elements of study
   - ① Use case-based investigation of “sufficient anonymization” and methods of producing anonymously processed information
   - ② Investigation of standards for security control measures for information on processing methods, etc.
   - ③ Relationships between “sufficient anonymization” and methods producing anonymously processed information
   - ④ Survey of acceptability to users
   - ⑤ Compilation of rules on location data handled by telecommunications carriers

3. Establishment of consultative council
   - A consultative council was established and met four times. The council consisted of five experts in fields including privacy of communications, privacy protection, anonymization technologies, and telecommunications policy, and included as observers the secretariat of the Personal Information Protection Commission of Japan, telecommunications carriers, and other interested bodies.

Council composition

*Members*

- Ryoji Mori* Attorney-at-law, Eichi Law Offices
- Shinsuke Ito Associate Professor, Faculty of Economics, Chuo University
- Ichiro Satoh Professor, National Institute of Informatics
- Katsumi Takahashi Executive Research Scientist, NTT Secure Platform Laboratories
- Toshiro Hikita Senior Researcher, Toyota InfoTechnology Center

*Study leader

*Observers*

- Secretariat of the Personal Information Protection Commission, Japan
- Telecommunications Carriers Association (TCA)
- Japan Data Communications Association
- NTT DOCOMO, Inc.
- KDDI Corp.
- Softbank Corp.
- NTT Broadband Platform, Inc.

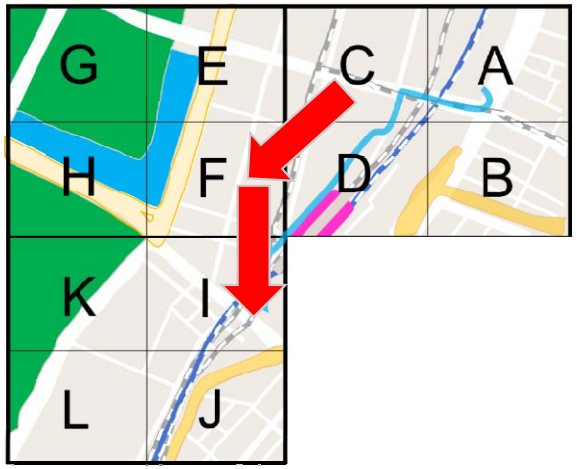
*Secretariat*

- Second Telecommunications Consumer Policy Division, Telecommunications Business Department, Telecommunications Bureau, Ministry of Internal Affairs and Communications (MIC)
- Nomura Research Institute, Ltd.

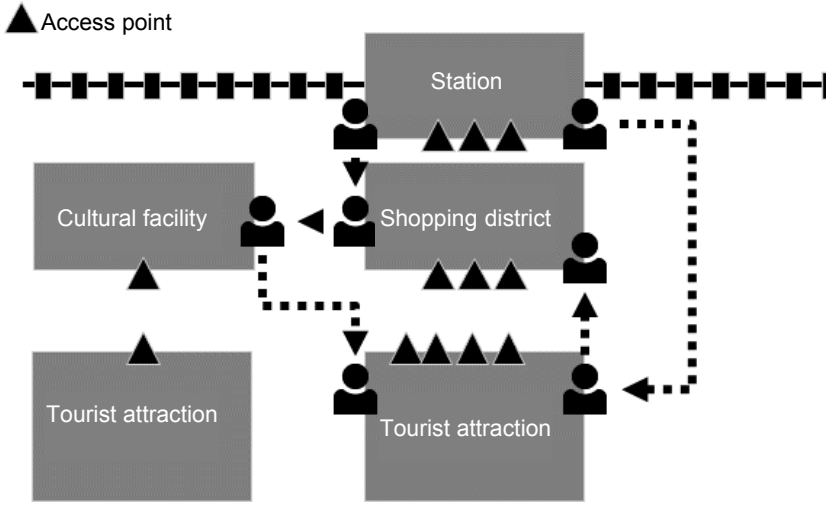
## 2. Study findings

# (1) Use case-based investigation of “sufficient anonymization” and methods of producing anonymously processed information: the use cases

- Commercial use case
  - Data assumed to be used for analysis of store market areas and sales promotion activities such as marketing support services.
  - Location data on a general office and commercial area of a city center visited by the general public, converted to 500 m, 250 m, and 125 m resolution grids to analyze movement between grid cells.
  - Analysis of location data supplemented by data on gender, age, municipality of residence, and also information on tastes and hobbies.




- Tourism use case
  - Data assumed to be used to determine movements of users (including tourists from overseas), routes used, etc.
  - Location data obtained from communications with access points installed at railway stations, tourist spots, etc. in tourism areas visited frequently by foreign tourists is aggregated at each point to analyze movement between points.
  - Analysis of location data supplemented by data on gender, age, municipality of residence, and also language data.



## 2. Study findings

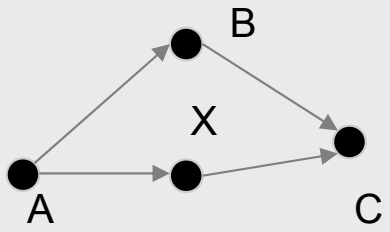
# (1) Use case-based investigation of “sufficient anonymization” and methods of producing anonymously processed information: key points regarding processing methods and identification of likely outputs

## “Sufficiently anonymized” information

*Key points*

- Data sets containing location data records and attributes aggregated to ensure non-identifiability.
- Assessed in terms of k-anonymity, k must be of a level that ensures non-identifiability under all circumstances.

*Likely outputs*




k individuals consisting of males aged 20-29 who moved A=>B=>C on holidays in December

| Attributes | Movement record | Number of records |
|---|---|---|
| Male Chiyoda resident aged 20-29 | A=>B=>C | 10 |
| Male Chiyoda resident aged 20-29 | A=>X=>C | 5 |
| … | … | .. |

## Anonymously processed information

*Key points*

- Data sets containing location data records and attributes processed so that specific individuals cannot be identified and original personal data cannot be restored.
- Produced by adjusting period covered by location data, location accuracy, and attribute granularity according to user needs (custom-made).
- Assessed in terms of k-anonymity, does not have to be processed so that k ≥ 2 depending on circumstances.

*Likely outputs*

| Pseudo ID | Attributes |
|---|---|
| 001 | Male Chiyoda resident aged 20-29 |
| 002 | Male Chiyoda resident aged 20-29 |
| … | … |

| Pseudo ID | Movement record |
|---|---|
| 001 | Date/time A⇒B⇒C |
| 001 | Date/time A⇒X⇒C |
| 001 | Date/time A⇒B⇒C |
| 002 | Date/time A⇒B⇒C |
| 002 | Date/time A⇒X⇒C |
| … | … |

Data sets tailored to user needs as long as unidentifiability is ensured.

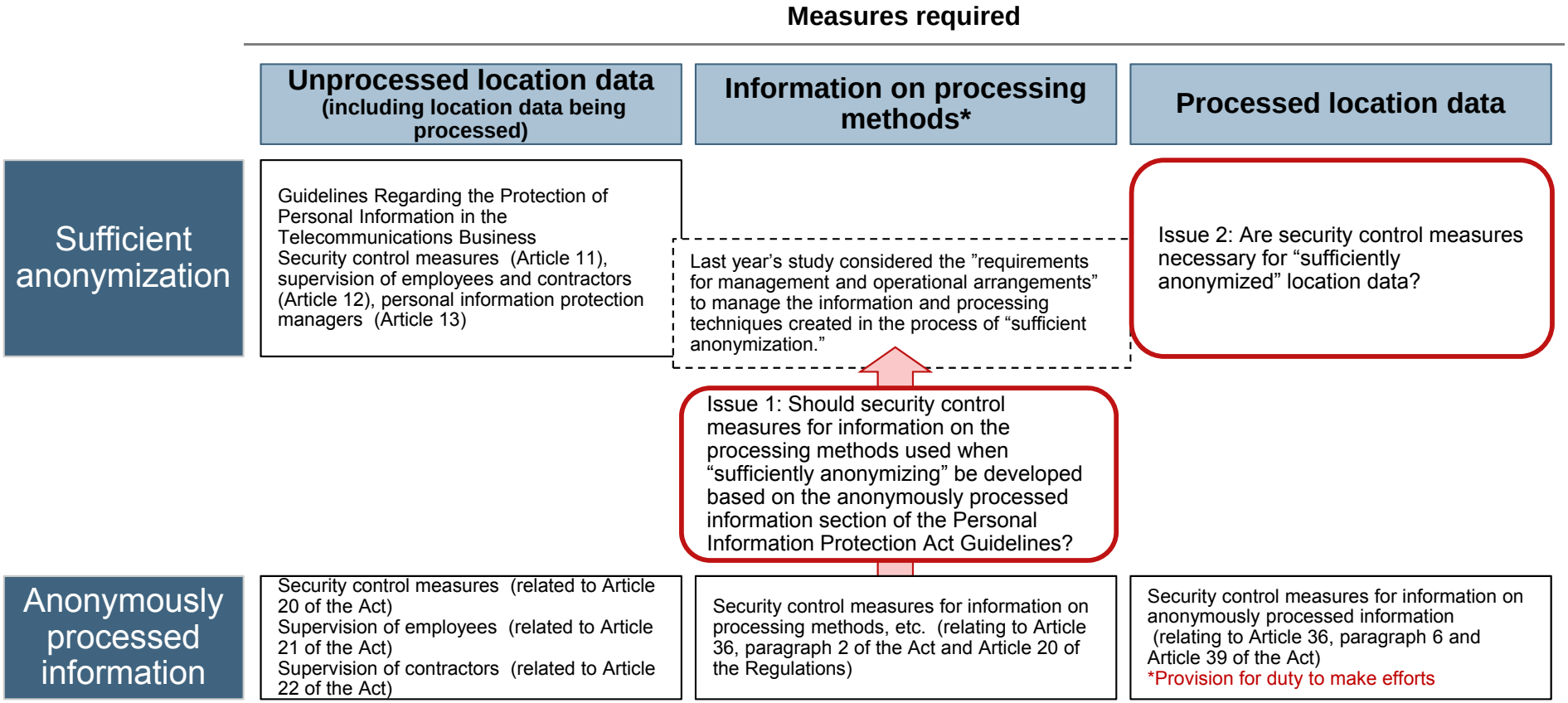
## 2. Study findings

# Evaluation factors and requirements pertaining to risk of identification of individual communications and specific individuals

| Evaluation factors | Requirements |
|---|---|
| 1) Supplementary information | • Data should be selected and processed while being mindful of the increased risk of identification of individual communications and specific individuals depending on the supplementary information used. |
| 2) Nature of location | • If location data contains information on individuals' homes and school or work commutation routes and destinations, it should be processed paying attention to this.<br>• If location data contains information on locations pertaining to sensitive personal information, it should be processed paying attention to this. |
| 3) Size of group | • If data concerns specific schools/workplaces or groups with unusual tastes of hobbies, there is an increased risk of identification of individual communications or specific individuals depending on the size of the group. Data should therefore be processed while being mindful of the size of the group. |
| 4) Characteristics of timing of acquisition | • As there is an increased risk that other information can be used to identify individual communications or specific individuals if the timing of acquisition coincides with the date or time of a particular event of incident, data should be processed while being mindful of the timing of its acquisition. |
| 5) Precision of location | • As highly precise location data poses a high risk of identification of individual communications or specific individuals, precision should be appropriately reduced. Particular attention should be paid in the case of data on areas of low population density. |
| 6) Period and scope of movement records | • Where movement records cover a long period or specific time of day, there is an increased risk concerning (a) patterning, (b) nature of location, and (c) identifiability. Data should therefore be processed while being mindful of this. |
| 7) Time accuracy and intervals | • When time of acquisition becomes more accurate or acquisition becomes more frequent, there is a heightened risk of identification of individual communications and specific individuals. Detailed time information can also function as a common identifier between different data sets when coupled with location data. Time accuracy should therefore be reduced and appropriate intervals provided between acquisition of data. |
| 8) Number of individuals covered | • Data should be processed while being mindful that the risk of identification of individual communications and specific individuals increases when the number of individuals contained in a data set is low. (The number of individuals should be counted as described in Appendix 4.)<br>• It should be borne in mind that there may be fewer individuals than mobile devices, as a single individual may carry more than one device. |
| 9) Period until provision of data | • If location data is provided as "sufficiently anonymized" information not long after the data was originally acquired, there is an increased risk that individual communications or specific individuals could be identified by consulting other information. Data should therefore be processed while being mindful of this. |

# (2) Investigation of standards for security control measures for information on processing methods, etc.: identification of issues

- When data is “sufficiently anonymized,” necessary and appropriate measures need to be taken to protect the privacy of communications and maintain the security of personal information (“security control measures”). Using the legal framework established for anonymously processed information by the amended Personal Information Protection Act as a reference point, the following issues were identified to assist the development of rules on such measures.




*Assuming use of hash function, etc. when pseudonymizing

(1) Articles of the Guidelines Regarding the Protection of Personal Information in the Telecommunications Business are numbered in accordance with the draft opinion document on revisions based on the amended Personal Information Protection Act that entered effect on January 19, 2017.
(2) Articles of the Personal Information Act are numbered in accordance with the act following its entry into effect on May 30, 2017.

# (2) Investigation of standards for security control measures for information on processing methods, etc.: findings on issues

| Issues | Findings |
|---|---|
| Issue 1<br>• Should security control measures for information on the processing methods used when “sufficiently anonymizing” be developed based on the anonymously processed information section of the Personal Information Protection Act Guidelines? | • The kinds of action that will be required to be taken as security control measures to protect information on the processing methods used to “sufficiently anonymize” data will be formulated using the security control measures for information on processing methods, etc. described in the anonymously processed information section of the Personal Information Protection Act Guidelines (relating to Article 36, paragraph 2 of the Act and article 20 of the Regulations) as a reference.<br>  • As there is a risk that information on processing methods could be used to identify specific individuals or individual communications from processed location data, security control measures should be of the same level as measures for unprocessed location data.<br>  • As location data in the process of being “sufficiently anonymized” is protected by the principle of secrecy of communications, it should be protected by security control measures equivalent to those required to protect secrecy of communications. |
| Issue 2<br>• Are security control measures necessary for “sufficiently anonymized” location data? | • As “sufficiently anonymized” location data presents a technologically considerably reduced risk of identification of specific individuals and individual communications, telecommunications carriers will not be required to implement special security control measures. |

# (3) Relationships between “sufficient anonymization” and methods of producing anonymously processed information, etc.: identification of issues

- Two issues were identified using a 2x2 matrix: along the vertical dimension, location data handled by telecommunications carriers is categorized according to whether it is protected (“private”) or not protected (“not private”) by the principle of secrecy of telecommunications; along the horizontal dimension, data is divided into “sufficiently anonymized” information (not limited to personal information) and anonymously processed information (corresponding to personal information).

| Location data (handled by telecommunications carriers) | Method of production, notification/consent, security control measures, etc. | |
|---|---|---|
| | “Sufficiently anonymized” information (not limited to personal information) | Anonymously processed information (corresponding to personal information |
| Private | Subject of last year’s study | Issue 1: Can location data that is protected by the principle of secrecy of communications even be used as anonymously processed information? |
| Not private | Issue 2: Should “sufficient anonymization” standards be required to be met? | |

# (3) Relationships between “sufficient anonymization” and methods of producing anonymously processed information, etc.: findings on issues

- Two issues were identified regarding the relationship of “sufficient anonymization” and methods of production of anonymously processed information, etc. These were considered from both institutional and technological angles, and the results are summarized below.

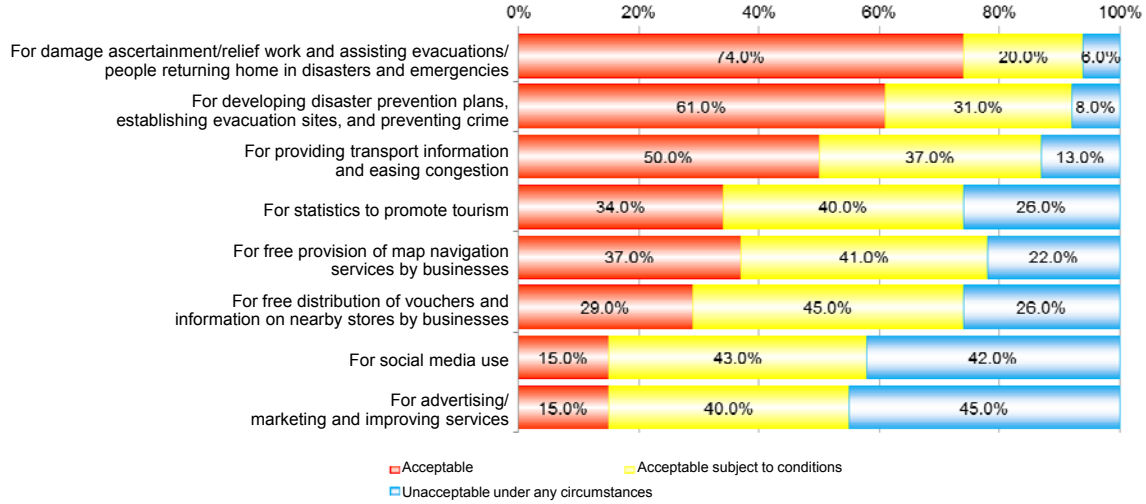
| Issues | Finding | Reasons |
|---|---|---|
| Issue 1 (private x anonymously processed information)<br>• Can location data that is protected by the principle of secrecy of communications even be used as anonymously processed information? | • Location data that is protected by the principle of secrecy of communications is not normally allowed to be processed into anonymously processed information and handled without valid consent. | • Institutional: When using data protected by the principle of secrecy of communications, valid consent is as a rule required as a means of ensuring the involvement of the individual concerned.<br>• Technological: As processing to a standard that prevents the identification of not only individuals but also individual communications is required to ensure protection of privacy of communications, in practice data needs to be anonymized in accordance with stricter standards than for anonymously processed information in order to protect personal information. |
| Issue 2 (not private)<br>• Should “sufficient anonymization” standards be required to be met when handling location data that is not protected by the principle of secrecy of communications? | • It is not necessary to require that “sufficient anonymization” standards be applied to anonymization of location data that is not protected by the principle of secrecy of communications.<br>• If, on the other hand, subscriber data or high-precision location data is used, processing to a standard equivalent to “sufficient anonymization” provides an effective and desirable means of protecting privacy. | • Institutional: The data protection objectives and regulations governing “sufficiently anonymized” data and anonymously processed information differ. Different frameworks should therefore be used according to whether the data concerned is protected by the principle of secrecy of communications or is classified as personal information, and there is no need to uniformly require all data to meet “sufficient anonymization” standards.<br>• Technological: “Sufficient anonymization” standards are in practice generally stricter than for processing of anonymously processed information. |

## 2. Study findings

# (4) Survey of acceptability to users: survey outline

- Survey objectives
  1. To determine whether “sufficient anonymization” and anonymously processed information differ in acceptability to ordinary users
  2. To determine what measures can be taken to ease concerns and improve acceptability to ordinary users
- Survey methodology
  - Combined use of central location test (CLT) and depth interviews
    - Anonymization methods are hard for ordinary users to grasp. So that appropriate responses could be obtained, therefore, it was decided that a CLT should be conducted, as this method allows background information and objectives to be explained in detail face to face.
    - The purpose of the depth interviews was to explore in greater depth the reasons for the responses to the CLT and changes in opinion when use conditions are altered.
- Number of participants
  - CLT: 100, depth interviews: 10
    - To avoid bias in the privacy sensitivity of CLT participants, pre-test questions were asked to gauge participants’ tolerance regarding use of location data. Participants were then recruited to ensure that the results obtained from the CLT participants as a whole approximated the responses to similar questions asked for a past MIC public survey.*

### Distribution of CLT participants’ responses (n=100)




### (Ref.) Distribution of responses to similar questions in the MIC Location Data Privacy Report (July 2014)

**Q. Which answer most resembles your opinion regarding use of your location data for the following purposes?**

(%)

| Type of use | | n= | Acceptable (total) | | | Acceptable | Acceptable subject to conditions | Unacceptable under any circumstances | Acceptable (total) |
|---|---|---|---|---|---|---|---|---|---|
| Type of use | Ascertainment of damage and relief work, assistance of evacuations and people returning home in disasters and emergencies | 1600 | | | | 67.3 | 28.1 | 4.6 | 95.4 |
| | Development of disaster prevention plans, establishment of evacuation sites, crime prevention | 1600 | | | | 58.3 | 35.5 | 6.2 | 93.8 |
| | Transport information and easing of congestion | 1600 | | | | 45.0 | 44.1 | 10.9 | 89.1 |
| | Statistics to promote tourism | 1600 | | | | 20.3 | 52.9 | 26.8 | 73.2 |
| | Free provision of map navigation services by businesses | 1600 | | | | 24.1 | 51.4 | 24.5 | 75.5 |
| | Free distribution of vouchers and information on nearby stores by businesses | 1600 | | | | 20.5 | 50.5 | 29.0 | 71.0 |
| | Social media use | 1600 | | | | 11.0 | 43.2 | 45.8 | 54.2 |
| | Advertising, marketing, and service improvement | 1600 | | | | 9.1 | 46.1 | 44.8 | 55.3 |

*Reference values (n < 30) shown in gray.

*Survey for the MIC Location Data Privacy Report (July 2014).

## 2. Study findings

# (4) Survey of acceptability to users: summary of results

Survey objective 1: To determine whether “sufficient anonymization” and anonymously processed information differ in acceptability to ordinary users

- Aggregation for “sufficient anonymization” is one factor that ordinary users feel makes it harder for individuals to be identified than with anonymously processed information.
  - A particularly marked difference is observed in case 2, where more items are processed by associating with subscriber data (see table below).
- Location data near users’ homes needs to be handled carefully.
  - In the case of both “sufficient anonymization” and anonymously processed information, approximately 60% of ordinary users who said that individuals could be identified considered place of residence (at the municipality level) to be most likely to lead to identification.
- Perceptions regarding the k value for “sufficient anonymization” were varied.
  - In the case of “sufficient anonymization,” some said that they would not be concerned if k = 2, while others called for k = 10 or higher.

Survey objective 2: To determine what measures can be taken to ease concerns and improve acceptability to ordinary users

- Providing opt-outs and restricting data recipients both help to ease ordinary users’ concerns.
  - Providing opt-outs regardless of use (90%) and restricting recipients (70%) help ease the concerns of ordinary users who are concerned about the provision of data to third parties.

**Results to questions that confirmed whether or not recipients thought that they could be personally identified from “sufficiently anonymized” information and anonymously processed information after being presented with specific processing methods based on use cases**

n = 100 in both cases

| Case 1: Tourism use | | Sufficient anonymization | |
|---|---|---|---|
| | | Not identifiable | Identifiable |
| Anonymously processed | Not identifiable | 43% | 10% |
| | Identifiable | 12% | 35% |

| Case 2: Commercial use | | Sufficient anonymization | |
|---|---|---|---|
| | | Not identifiable | Identifiable |
| Anonymously processed | Not identifiable | 38% | 6% |
| | Identifiable | 20% | 36% |

# (5) Compilation of rules on location data handled by telecommunications carriers

- Based on the results of investigation of (1)- (4), rules on location data were compiled into the following two categories. Appended information on specific use cases and approaches to them was also compiled.
  - Details that should be incorporated in rules on “sufficient anonymization”
  - Details that should be incorporated in rules on production of anonymously processed information using location data handled by telecommunications carriers

Details for incorporation into rules on “sufficient anonymization”

1. Purpose
2. Definition of terms
3. Scope of application
4. Handling of “sufficient anonymization”

4.1. Processing by “sufficient anonymization”
4.2. Security control measures
4.3. Notification, consent, and choice
4.4. Opt-outs
4.5. Privacy impact assessment (PIA)

5. Revision of rules

Appendix

Details for incorporation into rules on production of anonymously processed data using location handled by telecommunications carriers

1. Purpose
2. Definition of terms
3. Scope of applications
4. Standards on methods of production of anonymously processed information

4.1. Compliance with regulations relating to Article 36, paragraph 1 of the Personal Information Protection Act
4.2. Other measures for location data suited to the nature of distinctive personal information databases, etc.

5. Opt-outs
6. Revision of rules

Appendix

# 3. Future directions

1) Incorporation into industry group rules and personal information protection guidelines

- The present study resulted in the compilation of details for incorporation into rules on “sufficient anonymization” and rules on the production of anonymized processed information using location data handled by telecommunications carriers. It is anticipated that these will be taken up by industry groups and authorized personal information protection associations in the telecommunications business, which will use them as a model for when handling location data.
- The amended Personal Information Protection Act is scheduled to go into full effect on May 30, 2017. In conjunction with this, the “Guidelines on Personal Information Protection in the Telecommunications Business” will also enter effect. It is hoped that industry groups and authorized personal information protection associations will formulate voluntary rules and personal information protection guidelines based on the details compiled for this study, and MIC will need to organize initiatives to support these activities.

2) Follow-up of security control measures pertaining to “sufficient anonymization”

- Businesses are required to make efforts to implement security control measures to protect anonymously processed data even after it has been processed. As data that has been “sufficiently anonymized,” on the other hand, may be assumed (as the term suggests) to have been sufficiently anonymized, this study found that data that has been processed does not need to be protected by security control measures. However, advances in technology mean that it is hard to completely eliminate the possibility that individual communications and individuals could be identified from processed data, and so MIC will have to monitor the state of application of rules by businesses, and periodically revise them and take necessary steps in light of institutional changes and developments in technology.

3) Coordination with initiatives in other fields concerning anonymously processed information

- In the proposed rules on production of anonymously processed information compiled for the present study, it was suggested that data containing information on locations pertaining to sensitive personal information (as in the case of “nature of location”) should be processed paying attention to this. The commission regulations and guidelines established by the Personal Information Protection Commission, on the other hand, require that anonymously processed information be processed to a sufficient level to ensure that specific individuals cannot be identified, and do not go so far as to require that attention be paid to sensitive personal information. The rules compiled for this study therefore impose a stricter standard of processing than in general fields.
- Arrangements for anonymously processed information have only just been established and have not yet entered operation, and it is anticipated that both the authorities and businesses will engage in action by a process of trial and error. The rules compiled for this study too should be adjusted through trial and error based on initiatives in other fields.

4) Development of rules to accommodate development of IoT

- High-performance mobile devices such as smartphones with Bluetooth functionality allow location data to be captured by beacons too. It is expected that as IoT develops, various sensors will appear in everyday places and that various entities will gather location data by various means and use it.
- The process of acquisition, analysis, and use of such data involves not only the individual concerned and data user but numerous other stakeholders such as intermediaries and analysts, and it is essential that adjustments be made regarding, among other things, contractual relations between businesses and confirmation of the wishes of the individual. It will therefore be necessary to pursue the development of rules to enable the appropriate use of location data and other information that has a close bearing on privacy in light of the development of IoT.