International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

X.1148

(09/2020)

SERIES X: DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY

Secure applications and services (1) – Web security

Framework of de-identification process for telecommunication service providers

Recommendation ITU-T X.1148

ITU-T X-SERIES RECOMMENDATIONS

DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY

PUBLIC DATA NETWORKS	X.1–X.199
OPEN SYSTEMS INTERCONNECTION	X.200–X.299
INTERWORKING BETWEEN NETWORKS	X.300–X.399
MESSAGE HANDLING SYSTEMS	X.400–X.499
DIRECTORY	X.500–X.599
OSI NETWORKING AND SYSTEM ASPECTS	X.600–X.699
OSI MANAGEMENT	X.700–X.799
SECURITY	X.800–X.849
OSI APPLICATIONS	X.850–X.899
OPEN DISTRIBUTED PROCESSING	X.900–X.999
INFORMATION AND NETWORK SECURITY	
General security aspects	X.1000–X.1029
Network security	X.1030–X.1049
Security management	X.1050–X.1069
Telebiometrics	X.1080–X.1099
SECURE APPLICATIONS AND SERVICES (1)	
Multicast security	X.1100–X.1109
Home network security	X.1110–X.1119
Mobile security	X.1120–X.1139
Web security	**X.1140–X.1149**
Security protocols (1)	X.1150–X.1159
Peer-to-peer security	X.1160–X.1169
Networked ID security	X.1170–X.1179
IPTV security	X.1180–X.1199
CYBERSPACE SECURITY	
Cybersecurity	X.1200–X.1229
Countering spam	X.1230–X.1249
Identity management	X.1250–X.1279
SECURE APPLICATIONS AND SERVICES (2)	
Emergency communications	X.1300–X.1309
Ubiquitous sensor network security	X.1310–X.1319
Smart grid security	X.1330–X.1339
Certified mail	X.1340–X.1349
Internet of things (IoT) security	X.1360–X.1369
Intelligent transportation system (ITS) security	X.1370–X.1389
Distributed ledger technology security	X.1400–X.1429
Distributed ledger technology security	X.1430–X.1449
Security protocols (2)	X.1450–X.1459
CYBERSECURITY INFORMATION EXCHANGE	
Overview of cybersecurity	X.1500–X.1519
Vulnerability/state exchange	X.1520–X.1539
Event/incident/heuristics exchange	X.1540–X.1549
Exchange of policies	X.1550–X.1559
Heuristics and information request	X.1560–X.1569
Identification and discovery	X.1570–X.1579
Assured exchange	X.1580–X.1589
CLOUD COMPUTING SECURITY	
Overview of cloud computing security	X.1600–X.1601
Cloud computing security design	X.1602–X.1639
Cloud computing security best practices and guidelines	X.1640–X.1659
Cloud computing security implementation	X.1660–X.1679
Other cloud computing security	X.1680–X.1699
QUANTUM COMMUNICATION	
Terminologies	X.1700–X.1701
Quantum random number generator	X.1702–X.1709
Framework of QKDN security	X.1710–X.1711
Security design for QKDN	X.1712–X.1719
Security techniques for QKDN	X.1720–X.1729
DATA SECURITY	
Big Data Security	X.1750–X.1759
5G SECURITY	X.1800–X.1819

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T X.1148

Framework of de-identification process for telecommunication service providers

Summary

Telecommunication organizations collect, manage, use, and share data about individuals, including personally identifiable information. As a result, they utilize data de-identification techniques to protect individuals' data. Recommendation ITU-T X.1148 describes a framework of de-identification process with operational steps and specifies data release models and data stages in a de-identification process for telecommunication service providers based on data lifecycle model and the roles of stakeholders.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T X.1148	2020-09-03	17	11.1002/1000/14249

Keywords

Data principal, de-identification, de-identification process, k-anonymity, l-diversity, PII protection, release models, t-closeness.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Page

Introduction

With the rapid development of Internet based information and communication technologies and services, large amounts of data are generated, transmitted, and stored with explosive growth. Data are generated by many sources: not only sensors, cameras or network devices, but also web pages, email systems or social networks, and many others. Datasets are becoming so large, complex and are arriving so fast that traditional data process methods and tools are no longer adequate. Efficient data analytics within tolerable delay becomes very challenging. A paradigm called big data analytics are being developed to resolve the above issues.

Telecommunication organizations collect, manage, use, and share data about individuals, including personally identifiable information. As a result, they utilize data de-identification techniques to protect individuals' data. The relationships between parties participating in the data flow for data exchange affect whether data de-identification needs to be performed before its collection, after its collection but before its storage, or only before it is shared with the next party in data exchange. Accordingly, telecommunication service providers need to provide data de-identification as a service in a timely, efficient, and safe way to data customers.

Recommendation ITU-T X.1148

Framework of de-identification process for telecommunication service providers

1 Scope

This Recommendation provides an overview of de-identification process based on data lifecycle model, specifies a de-identification process framework with operational steps and roles of stakeholders in the de-identification process. It further discusses data release models and data stages in a de-identification process and includes various de-identification approaches and examples in its annexes and appendices.

This Recommendation does not address issues related to regulation.

2 References

None.

3 Terms and definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 aggregated data [b-ISO/IEC 20889]: Data representing a group of data principals, such as a collection of statistical properties of that group.

3.1.2 anonymization [b-ISO/IEC 29100]: Process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party.

3.1.3 attribute [b-ISO/IEC 20889]: Inherent characteristic.

3.1.4 dataset [b-ISO/IEC 20889]: Collection of data.

3.1.5 de-identification [b-ISO 25237]: General term for any process of reducing the association between a set of identifying data and the data subject (see clause 3.2.4).

3.1.6 de-identification process [b-ISO/IEC 20889]: Process of removing the association between a set of identifying attributes and the data principal.

3.1.7 de-identification technique [b-ISO/IEC 20889]: Method for transforming a dataset with the objective of reducing the extent to which information is able to be associated with individual data principals.

3.1.8 de-identified dataset [b-ISO/IEC 20889]: Dataset resulting from the application of a de-identification process.

3.1.9 de-identified information [b-NISTIR 8053]: Record that have had enough PII removed or obscured such that the remaining information does not identify an individual and there is no reasonable basis to believe that the information can be used to identify an individual.

3.1.10 differential privacy [b-ISO/IEC 20889]: Formal privacy measurement model that ensures that the probability distribution of the output from a statistical analysis differs by at most a specified value, whether or not any particular data principal is represented in the input dataset.

NOTE − More specifically, differential privacy provides:

a) a mathematical definition of privacy which posits that, for the outcome of any statistical analysis to be considered privacy-preserving, the analysis results from the original dataset are indistinguishable from those obtained if any data principal is added to or removed from the dataset; and

b) a measure of privacy that enables monitoring of cumulative privacy loss and setting of an upper bound (or "budget") for loss limit. A formal definition is as follows. Let ε be a positive real number, and M be a randomized algorithm that takes a dataset as input. The algorithm M is said to be ε-differentially private if for all datasets D1 and D2 that differ in a single element (i.e., the data for one data principal), and all subsets S of the range of M, mml_m1, where the probability is taken over the randomness used by the algorithm.

3.1.11 identifier [b-ISO/IEC 20889]: Set of attributes in a dataset that enables unique identification of a data principal within a specific operational context.

NOTE − See Annex B for a discussion of how this definition relates to those given in other standards.

3.1.12 identifying attribute [b-ISO/IEC 20889]: Attribute in a dataset that can contribute to uniquely identifying a data principal within a specific operational context.

3.1.13 privacy stakeholder [b-ISO/IEC 29100]: Natural or legal person, public authority, agency or any other body that can affect, be affected by, or perceive themselves to be affected by a decision or activity related to personally identifiable information (PII) processing.

3.1.14 pseudonymization [b-ISO/IEC 20889]: de-identification technique that replaces an identifier (or identifiers) for a data principal with a pseudonym in order to hide the identity of that data principal.

3.1.15 quasi-identifier [b-ISO/IEC 20889]: Attribute in a dataset that, when considered in conjunction with other attributes in the dataset, singles out a data principal.

3.1.16 record [b-ISO/IEC 20889]: Set of attributes concerning a single data principal.

3.1.17 re-identification [b-ISO/IEC 20889]: Process of associating data in a de-identified dataset with the original data principal.

NOTE − A process that establishes the presence of a particular data principal in a dataset is included in this definition.

3.1.18 single out [b-ISO/IEC 20889]: Isolate records belonging to a data principal in the dataset by observing a set of characteristics known to uniquely identify this data principal.

3.1.19 third party [b-ISO/IEC 29100]: Privacy stakeholder other than the personally identifiable information (PII) principal, the PII controller and the PII processor, and the natural persons who are authorized to process the data under the direct authority of the PII controller or the PII processor.

3.1.20 Trusted Third Party [b-ISO/IEC 18014-1]: Security authority, or its agent, trusted by other entities with respect to security related activities.

3.1.21 k-anonymity [b-ISO/IEC 20889]: Formal privacy measurement model that ensures that for each identifier in a dataset there is a corresponding equivalence class containing at least K records.

3.1.22 l-diversity [b-ISO/IEC 20889]: Formal privacy measurement model that ensures that for a selected attribute each equivalence class has at least L well-represented values.

NOTE − L-diversity is a property of a dataset that gives a guaranteed lower bound, L, on the diversity of values shared by an equivalence class for a selected attribute.

3.1.23 t-closeness [b-ISO/IEC 20889]: Formal privacy measurement model that ensures that the distance between the distribution of a selected attribute in an equivalence class and the distribution of this attribute in the entire table is no more than a threshold T

NOTE − A table is said to have T-closeness with respect to a selected attribute if all equivalence classes containing this attribute have T-closeness.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 data controller: Stakeholder (or privacy stakeholder) that determines the purposes and means for processing data other than natural persons who use data for personal purposes.

3.2.2 data processor: Stakeholder that processes data on behalf of and in accordance with the instructions of a data controller.

3.2.3 data protection officer: Person appointed by the personally identifiable information (PII) controller to ensure, in an independent manner, compliance with the privacy law or regulation requirements.

NOTE − "PII controller" is synonym of "data controller".

3.2.4 data subject: Entity to which data relates.

NOTE − "data subject" is synonym of "PII principal" and "data principal".

3.2.5 process: In relation to information or data, this means obtaining, recording, or holding the information or data or carrying out any operation or set of operations on the information or data, including:

− organization, adaptation or alteration of the information or data,

− retrieval, consultation or use of the information or data,

− disclosure of the information or data by transmission, dissemination or otherwise making available, or

− alignment, combination, blocking, erasure or destruction of the information or data.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

DP	Differential Privacy
DPO	Data Protection Officer
PII	Personally Identifiable Information
TTP	Trusted Third Party

5 Conventions

None.

6 Overview of de-identification process

The purpose of de-identification process is to protect the confidentiality of a subjects' data. As these data may include personally identifiable information (PII), before and after data analytics with the purpose of extracting meaningful information, a data analyst must include security considerations.

This clause defines data analysis environments, data lifecycle model, roles of entities in the de-identification process and other de-identification considerations.

6.1 Data lifecycle model and de-identification phase

Typically, an organization sets de-identification objectives for privacy and security aims. This clause defines a data life cycle and describes when to consider a de-identification process based on this data life cycle model.

The data lifecycle concept is used to select the appropriate controls based on the analysis of the possibility of re-identification. This Recommendation defines a data life cycle as outlined in clauses 6.1.1 to 6.1.5.

Figure 1 provides an overview of the de-identification process in the data lifecycle model.

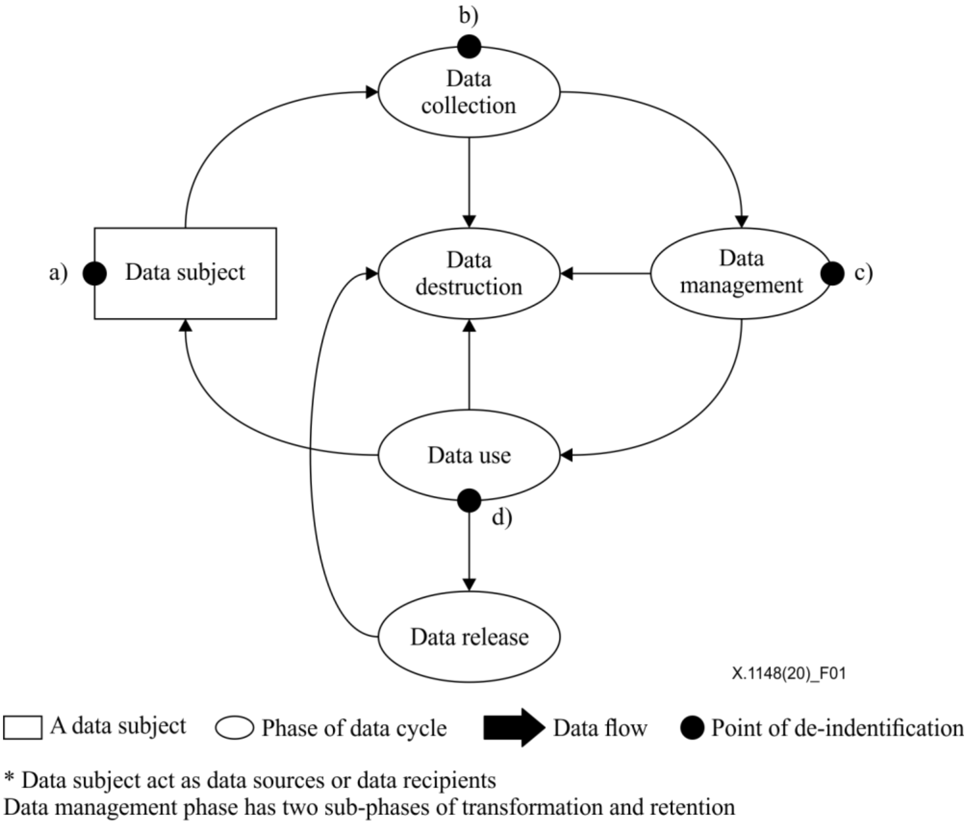

b)
Data collection
a)
Data subject
Data destruction
Data management
c)
Data use
d)
Data release
X.1148(20)_F01
A data subject
Phase of data cycle
Data flow
Point of de-indentification
* Data subject act as data sources or data recipients
Data management phase has two sub-phases of transformation and retention

Figure 1 – De-identification process in data lifecycle model

6.1.1 Data collection phase

Data are collected from data subjects which are persons to whom the data refers. The dataset produced as a result of this data collection may include PII. De-identification creates a new dataset from which all PII has been removed. It is recommended that de-identified datasets are internally used by an organization instead of the original dataset, wherever possible.

Using this model, de-identification can take place either:

- during data collection, i.e., (b) in Figure 1; or
- in the event that data were collected but identifier was not actually needed, i.e., (a) in Figure 1.

Identifiers that are not needed for data management (data transformation and data retention) should not be collected.

6.1.2 Data management phase

To avoid archiving the identifier, de-identification should be applied after data transformation and prior to data retention, i.e., (c) in Figure 1. Organizations are recommended to consider the potential for re-identification and set clear access controls, maximum retention limits, and data removal policies that maximally reduce the potential for linkability between de-identified data. Organizations are recommended to consider anonymization techniques such as data aggregation wherever the intended purpose of use allows.

6.1.3 Data use phase

If PII is needed within an organization for data management, the data is recommended to be de-identified prior to being released as a dataset for data sharing, i.e., (d) in Figure 1.

6.1.4 Data release phase

Data can be shared with third parties who are bound by additional administrative controls such as 'data-sharing' agreements. De-identified datasets may also be subject to release. De-identified data release is classified into three models: public, semi-public, or non-public. The amount of de-identification required may vary depending on the release model selected.

6.1.5 Data destruction phase

Data destruction can be done at any phase, i.e., data collection, data management, data use, and data release. Data should be destroyed with verified measures to avoid recovery of data. In particular, destruction of data should be considered upon detection of the possibility of re-identification.

6.2 De-identification considerations

Applying de-identification throughout the data life cycle increases its effectiveness. However, the nature of relationships among the parties participating in the data flow affects whether data de-identification needs to be performed before its collection, i.e., (a) in Figure 1, after its collection, i.e., (b) in Figure 1, but before its retention, i.e., (c) in Figure 1, or only before it is shared with the next party in the data flow, i.e., (d) in Figure 1. This decision, in turn, affects the feasibility of security and other organizational measures to enhance the effectiveness of a particular de-identification technique in each use case. Although de-identification may be a useful technique to protect the confidentiality of a subjects' data in cases where the purpose of use does not support anonymization techniques, it is not in itself sufficient to protect the subjects' data and must be considered as part of a comprehensive data protection framework. This clause describes the features and considerations of each phase.

6.2.1 Data collection

Local de-identification (or de-identification at the source) is the most prominent approach, which allows an individual (or a controller processing data for an individual) to remove all PII prior to releasing the data for analytics.

One de-identification aspect directly related to the data collection phase is data minimization. Each data controller who is collecting a subjects' data is required to precisely define what data is strictly necessary for the intended purpose of use, and limit data collection to only those defined parameters.

Specific processes should be in place to exclude unnecessary PII from data collection or transfer, in order to reduce data fields.

Another de-identification aspect is data aggregation. Data controllers are required to consider the aggregation of data in all cases wherein the purpose of use does not strictly require singling out individual data subjects.

6.2.2 Data management

6.2.2.1 Data transformation

Data transformation phase may include the application of de-identification techniques such as aggregation, statistical disclosure limitation, encryption, etc. Data transformation may be applied at one or multiple stages, including directly after collection and prior to long-term retention, after a substantial retention period and prior to access, or integrated with access.

The common transformation of data redacting or aggregating can be employed any time after collection and up until release. If applied immediately after collection, redacting or aggregating data may reduce the potential harm to data subjects in case of a data breach; however, doing so also curtails the potential to link, merge, or update the data after redaction.

The choice of data transformation method should be made after careful consideration of the potential harm of exposure to data subjects. The transformation decision should also take into account the analyses that must be supported by the purpose of data use later, as the techniques employed for reducing disclosure risks can affect the potential for later uses and analyses.

6.2.2.2 Data retention

Data retention is described as the process of storing data, including PII, to any form of non-volatile storage by a data controller or a party acting under the controller's direction. Information security and privacy controls already focus on the retention phase; thus, this clause summarizes controls without providing detailed considerations [b-ISO/IEC 27001]. A number of information security and privacy controls are common at the retention stage, such as access control, maintenance, security assessments, authentication procedures, incident monitoring and response, and audits.

In particular, organizations should follow maximum data retention and removal polices to ensure that data are retained for no longer than what is strictly necessary to fulfil the purpose of use, and that data is completely destroyed after this maximum retention period. For example, data sharing agreements often specify that the recipient must destroy the data within a specified period of time, such as one year after receipt, and laws may also require such a contractual provision.

6.2.3 Data use

De-identified data can be collected, stored, or shared for a range of purposes and applications, each relying on certain data properties being preserved after de-identification. One of the main reasons for releasing de-identified datasets is to provide others with an opportunity to study the values and properties of the raw data for research purposes [b-ISO/IEC 20889]. De-identification, therefore, should also seek to preserve as much utility in the information as possible, while protecting the privacy of individuals. This dual purpose of de-identification makes it an important approach to consider for use in a number of contexts, including data release models.

When releasing de-identified data, an organization must make a decision, typically by an expert committee that includes a broad range of stakeholders, to consider the potential impacts on data subjects related to release. Risk assessments and checklists are often used to guide this evaluation and determine an appropriate release mechanism that mitigates re-identification risk.

The choice of de-identification techniques depends on the degree of their applicability to, or 'utility' in, a particular use case.

7 De-identification process framework

This clause describes a de-identification process framework to provide de-identified PII in four steps, as shown in Figure 2 [b-KOREA].

Step 1 – Preliminary review

Step 1 involves verifying whether the target data is PII or not. If the data does include PII, proceed to Step 2. De-identification is needed.

Step 2 – De-identification

Step 2 involves de-identifying data to prevent inferences of specific individual information from the target dataset. This step invokes methods to remove or transform PII elements, either in total or partially. PII elements include identifiers, quasi-identifiers, and sensitive attributes.

Step 3 – Adequacy assessment

Step 3 involves assessing the adequacy of the de-identified dataset including PII elements. Relevant considerations include whether the target dataset still contains PII, direct potential of re-identification, potential of linkability that can lead to re-identification.

Step 4 – Follow-up management

Step 4 involves measuring managerial and technical safety to prevent re-identification.

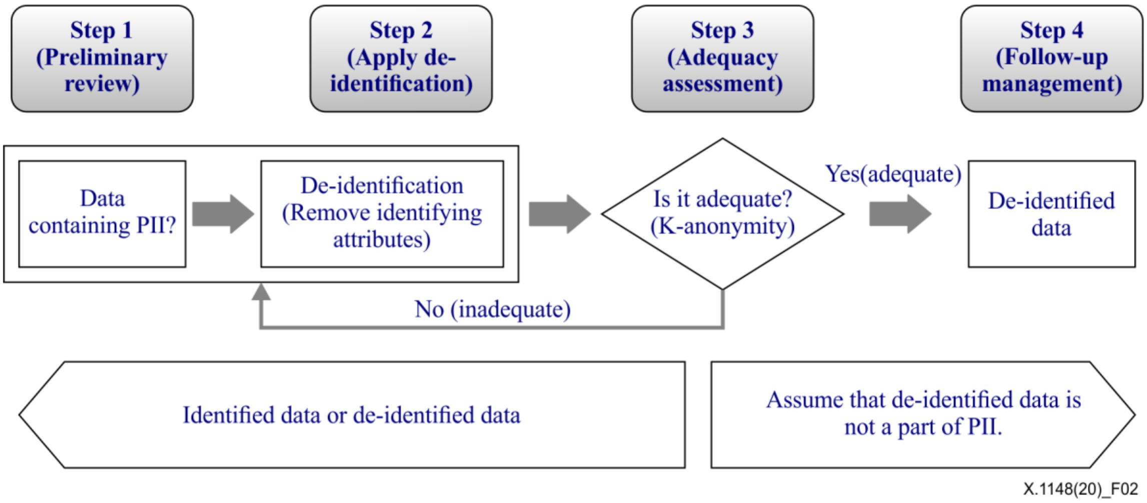

Step 1 (Preliminary review)
Step 2 (Apply de-identification)
Step 3 (Adequacy assessment)
Step 4 (Follow-up management)
Data containing PII?
De-identification (Remove identifying attributes)
Is it adequate? (K-anonymity)
Yes(adequate)
De-identified data
No (inadequate)
Identified data or de-identified data
Assume that de-identified data is not a part of PII.
X.1148(20)_F02

Figure 2 – De-identification process

Each of these steps is further described in clauses 7.1 to 7.4.

7.1 Step 1 – Preliminary review

Organizations that intend to use or provide data for various purposes should first determine their policies and standards. It is recommended that policies and standards include the following:

- what is the purpose and intended use for the de-identified information?
- what kind of data attributes does de-identified data consist of?
- what techniques are used for de-identification?
- what are the risk levels and adverse effects of re-identification?
- what solutions are available if a specific individual is re-identified?
- how is the level of re-identification evaluated?
- how is the manpower and cost for de-identification determined?

The specific considerations that constitute the preliminary review may vary depending on the type of data and the intended purpose of use. However, it is recommended that a set of standards be established.

Organizations that intend to process data for a number of purposes shall refer to the appropriate standards to verify whether specific data are PII or not. Even if data is not determined as PII, the organization is required to consider any potential for linkability between available data and take appropriate measures to minimize this risk. If yes, the de-identification step is necessary.

Examples of judgment criteria for PII include:

- there are no special limitations on data regarding its type, form, characteristics and format;
- if a data controller can identify an individual using the data, such data is considered a PII;

- data has to be about an individual. A statistical value of a group that consists of multiple individuals is not PII;
- data which may identify an individual via combination with an additional information is considered a PII. Additional information normally refers to publicly or easily available information.

7.2 Step 2 – Applying de-identification

7.2.1 De-identification for identifiers

An "identifier" is data such as a value or a name uniquely assigned to an individual or a thing that is related to an individual. In general, the collection of "identifiers" should be minimized, and any identifiers included in datasets should be deleted.

However, an identifier that is strictly necessary to the intended purpose may include data such as:

- unique identifier (resident registration number, social security number (SSN), passport number, foreigner's ID number, driver's license number, etc.);
- name (in Chinese characters, English name, etc.);
- detailed address (house number, street addresses, etc.);
- date (birth date, anniversary (wedding, etc.), certificate date, etc.);
- phone number (mobile, home, office, fax, etc.);
- medical record number, national health insurance number, welfare recipient number, etc.;
- bank account number, credit card number, etc.;
- photos (still picture, video, closed circuit television (CCTV) video, etc.);
- biometric data (fingerprint, voice, iris, etc.);
- e-mail address, IP address, media access control (MAC) address, homepage uniform resource locator (URL), etc.;
- identification code (employee number, customer number, etc.);
- other unique identification number (military service number, business registration number, etc.).

7.2.2 De-identification for quasi-identifier and high-identifiability attributes

In general, quasi-identifiers included in datasets should be removed if they are irrelevant to the purpose for which the data is used. De-identification techniques such as pseudonymization and aggregation should be applied if a quasi-identifier related to the use of data has identifiable elements.

Data that carries a high potential for identifiability, such as behavioral information, must be subject to de-identification and, wherever possible, anonymization techniques.

7.2.3 De-identification techniques

A range of techniques including pseudonymization, aggregation, data suppression and data masking can be used individually or in combination. Applying a pseudonymization technique alone may not be a sufficient technique of de-identification.

Various types of techniques are readily available to realize each technique. The most suitable technique should be chosen and utilized based on the purpose of data usage, and the strengths and weaknesses of each specific technique. Once de-identification is completed, one can proceed to the next step.

7.3 Step 3 – Adequacy assessment for de-identification process

An individual could be identified by combining other data or using various inference techniques when de-identification is not sufficient.

To reduce re-identification risk, the adequacy assessment of de-identified data is necessary before use. This includes assessing questions such as:

- what is the purpose of this de-identification request?
- what type of data attributes are involved in de-identification (including identifiers, or not)?
- what is the appropriate level of de-identification?

This adequacy assessment could be performed by a data protection officer (DPO), a commissioned trusted third party (TTP) or by an external assessment panel.

K-anonymity model is used among other privacy protection models when assessing adequacy. The k-anonymity model is a basic means of assessment. Additional assessment models (l-diversity, t-closeness, differential privacy (DP), etc.) may be applied if necessary.

Refer to Annex A for additional details on adequacy assessment.

7.4 Step 4 – Follow-up management

7.4.1 Protection measures for de-identified data

Protection measures are implemented to prevent the possibility of re-identifying de-identified data if it is leaked and/or combined with other data. These include measures such as:

- managerial protection measures: designate a person who will be in charge of the de-identified data files, determine the sharing of de-identified data, and destroy the data once its purpose of usage is achieved;
- technical protection measures: restrict access to de-identified data files, manage access records, and install and operate security programs.

Additionally, security measures also include protective measures to be taken in the event that de-identified data are leaked. These include measures such as:

- analyzing the cause of leakage and implementing both managerial and technical safety measures to prevent additional leakage;
- withdrawing and destroying leaked de-identified data.

7.4.2 Monitoring possibilities of re-identification

The data controller who intends to use de-identified data or provide it to a third party shall regularly monitor the possibilities of re-identification.

Upon detection of possibility of re-identification, suspension of processing, withdrawal and destruction of data must be requested to the data controller who has been provided the de-identified data.

7.4.3 Requirements for contract to third party

Re-identification risk management shall be included in a contract when providing or entrusting de-identified data to a third party for their usage. Re-identification risk management includes:

- notification to data subjects of data disclosure to third parties;
- provision of anonymized data to third parties wherever possible;
- prohibition of re-identification: stipulate that data controller which is given or consigned to process de-identification data shall be prohibited to re-identify the data by combining it with other data;

– restriction on re-provision or re-entrustment: stipulate the permitted scope of re-provision or re-entrustment in a contract when providing or entrusting processing of de-identified data;

– notification on risks of re-identification: stipulate the obligation to stop data processing and to inform the consignor and the consignee of the re-identification issue when the data are re-identified or the re-identification possibility becomes high.

7.4.4 Countermeasures to re-identification

In the event that de-identified data are re-identified, processing of data should be stopped, and the necessary measures should be taken to prevent leakage of PII.

Re-identified data shall be immediately destroyed.

8 De-identified data utility

8.1 De-identified data stages

This clause defines data de-identification stages that can be represented as data types to describe the degree to which an individual is directly identified by the data and how the individual is associated with characteristics (attributes) in the data. The specification of data in the context of data use or data processing should include not only the type of data, but also a description of the degree to which the data can identify an individual or associate an individual with a set of characteristics in the data.

Figure 3 provides data stages from identified data to de-identified data in the de-identification process. Each stage has a different possibility of re-identification risk as a spectrum. A data type characterizes specific stages that a dataset would go through as it is increasingly de-identified.

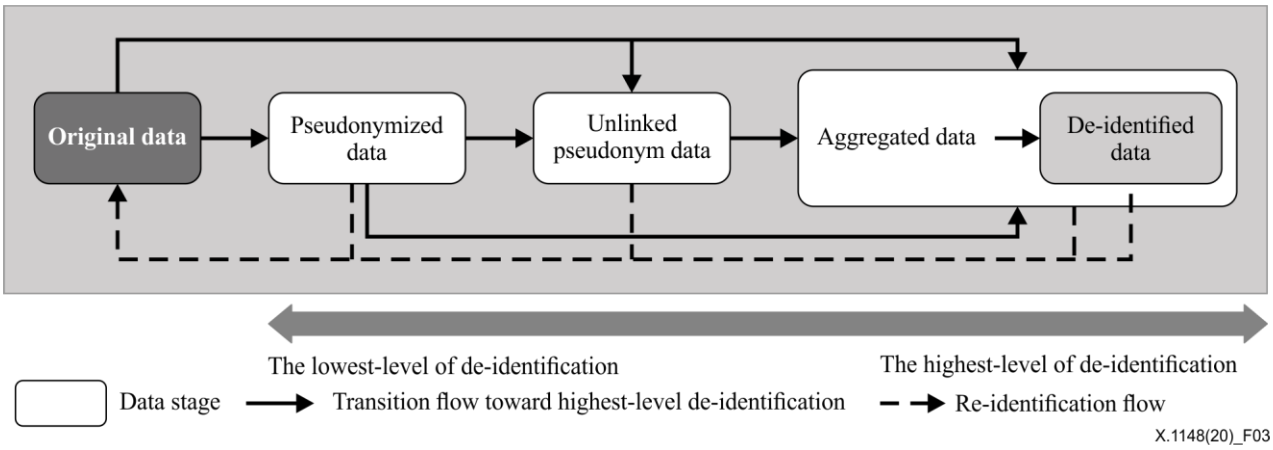


Figure 3 – De-identified data stages

As shown in Figure 3, all data exist on the de-identification stage. At the right (the highest-level de-identification) are de-identified data that are not related to individuals (for example, historical weather records) and therefore seek no privacy risk. At the left end (the lowest-level de-identification) are identified data that are linked directly to specific individuals. Between these two data stages are data that can be linked with effort, that can only be linked to groups of people, and that are based on individuals but cannot be linked back to them. In general, de-identification processes are designed to push data to the right while retaining some desired utility, lowering the risk of distributing de-identified data to a broader population or the general public.

8.1.1 Original data stage

In the original stage of identified data, data can unambiguously be associated with a specific person because an individual is observable in the information. Guidance on what can be considered as identifiers can be found in clause 4.4.1 of [b-ISO/IEC 29100].

8.1.2 Pseudonymized data stage

In the pseudonymized data stage, data cannot be reversed by reasonable efforts of anyone other than the party that performed the alias assignment because all identifiers are substituted by aliases. However, pseudonymized data may still be re-identified through linkability with other data.

This corresponds to data defined as "pseudonymization" in clause 3.1.14.

8.1.3 Unlinked pseudonymized data stage

In the unlinked pseudonymized data stage, all identifiers are erased or substituted by aliases for which the assignment function is erased or irreversible, such that a linkage cannot be re-established by reasonable efforts by anyone including the party that performed them. However, unlinked pseudonymized data may still be re-identified through linkability with other data.

8.1.4 Aggregated data stage

In the stage of aggregated data, data forms information about enough different persons that individual-level attributes are not inferable as statistical data that does not contain individual-level entries and is combined. Using aggregation techniques, all aggregated data does not reach the degree of identifiability below a threshold if cell size for a given crossing of some combination of variables can lead someone to identify a particular individual.

This corresponds to data defined as "aggregated data" in clause 3.1.1.

8.1.5 De-identified data stage

In the de-identified data stage, data is unlinked, and attributes are altered (e.g., attributes' values are randomized or generalized) in such a way that there is a reasonable level of confidence that a person cannot be identified, directly or indirectly, by the data alone or in combination with other data.

8.2 Data release models

De-identified data release model is classified into three models according to the situation of data analytic contexts [b-UKAN].

There are three release models for delivering de-identified data: public, semi-public, or non-public.

Each release model allows for different levels of availability and protection of information. Depending on the purposes and/or legislative requirements of the data release, the suitability of each model may vary. The release model plays an important role in the de-identification process, as the amount of de-identification required may vary depending on the release model selected.

Each of the three release models is reviewed in clauses 8.2.1 to 8.2.3.

8.2.1 Public data release model

In a traditional public data release, anyone can access the data without registration or conditions. Examples of such releases include the publicly available data from organizations and data posted to open-access data repository such as a web portal. Organizations proactively release datasets and make them freely available to anyone for use and republishing.

When releasing data publicly, it is common practice to place as few restrictions as possible on the information, including who can access it and how. As such, when individuals who download the dataset cannot be identified, these disclosures should be handled as public data releases.

Although in case of access to information requests in clause 8.2.2, it should be handled as public data releases in cases that do not require the person requesting information to agree to terms or conditions regarding the processing, privacy, or security of the information.

8.2.2 Semi-public data release model

The semi-public data sharing model is more restrictive than the public data release model and occurs when there exists a formal request and approval process to obtain access to data. In this case, the data recipient may agree to some terms of use or sign a "click-through" contract. Click-through contracts are online terms of use that may place restrictions on what can be done with the data and how the data are handled. Regardless, anyone can still download such data.

De-identification may also be useful in responding to access for information requests for datasets. By using de-identification, organizations can respond to requests in a privacy-protective manner while preserving the utility of the information. Organizations can use access controls for some restrictions when sharing data through an information system such as:

- requiring all users to register and provide contact information before accessing the data;
- employing authentication protocols to verify the identity of an individual;
- using tiered access systems to grant different levels of access to different parties based on, for example, affiliations or credentials of the individual.

With such information systems, an interactive query system might be made available to a community of researchers, and raw data might be made available to a small number of analysts who are approved through a careful screening process.

Also, the case of data access that does not require any data sharing occurs when analysts request that the data controller perform an analysis on their behalf. Therefore, this case might not involve the sharing of data by the organization.

8.2.3 Non-public data release model

Datasets that contain PII may be shared within and among organizations only if the disclosure is permitted under country regulatory guidance. If the disclosure is not permitted and the institutions still wish to share datasets, then any PII must be removed. Non-public data releases provide the least availability, but a higher amount of protection, requiring a smaller amount of de-identification.

When sharing information among organizations, because access to the dataset is limited to the organization, requirements regarding the privacy and security of the information can be set and enforced through a data-sharing agreement. For a data release to be treated as non-public, there must be a data-sharing agreement in place between the parties. The data-sharing agreement is an important part of a risk mitigation strategy in these releases, which includes some common terms such as:

- specification of those permitted access (recipient controls);
- data security requirements (infrastructure controls);
- restrictions on use, particularly prohibition against linking with other files and on deliberate re-identification (other data and governance controls);
- requirement to destroy the data once the use is complete (governance controls).

The purpose of data sharing agreement is threefold:

- it clearly distinguishes between those individuals or organizations the data controller trusts and those that it does not;
- it is a framework that specifies conditions under which access can occur;
- it can specify sanctions or penalties in case that the individual/organization transgress on those access conditions.

8.2.4 Comparison of data release models

In a data flow environment, one way to limit the chance of re-identification is to place controls on the way that the data may be obtained and used. These controls can be classified according to different data-release models, each with different advantages and risks. Organizations may also choose to apply a tiered access approach that combines several of these models to address a variety of use cases and privacy threats. Additionally, release models should consider the possibility of multiple or periodic releases. Several named models are ranged from no restrictions to tight restrictions. Table 2 provides a comparison of the data release models.

Table 2 – Comparison of data release models

	Public release model	**Semi-public release model**	**Non-public release model**
Access rights	• Everyone has access to released data freely	• Access to released data (or subset) from restrictive individual or organization	• Access to released data is to a subset of individuals or organizations
Use cases	• Unrestricted data access through web portal. i.e., freely available to anyone	• On-site safe setting • Delivered access • Virtual access remotely • Access through analysis server	• Sharing within and among organizations
Rights	• Unlimited rights to reuse and redistribute data	• Available to authorized individual or organization	• Re-use, republishing, or distribution of data is forbidden
Re-identification attack	• A demonstration attack for publicity	• Deliberate insider attack • Inadvertent recognition of an individual in the dataset by an acquaintance • Data leakage	

8.3 Relation between data release model and data stage

8.3.1 Non-public data release model

When sharing data from a data source to the non-public release model, the data requires de-identification. Under normal circumstances, in spite of the non-public release model, unlinked pseudonymized data and higher-level de-identification data would be used. In this case, the de-identification tools such as pseudonymization, cryptographic, synthetic, suppression, etc., can be used.

However, if there is a special contract or law between two sides, then the pseudonymized data could be used for analysing and storing data during this phase.

8.3.2 Semi-public data release model

When sharing data from a data source to the semi-public release model, it needs a higher level of de-identification than the non-public release model. It performs statistical processing to prohibit re-identification. After this, aggregated data and higher-level de-identification data could be released to the semi-public release model. More specifically, de-identification tools such as statistical, randomization, etc., can be used.

As shown in Table 2, a relatively lower level of de-identification can be allowed than with the public release model, since only restrictive individuals or organization can access the data.

8.3.3 Public data release model

When sharing data from a data source to the public release model, it needs a higher level of de-identification than for the semi-public release model. It performs the process to obtain de-identification data and after this process, the results can be used for the public release model, as shown in Table 2.

Annex A

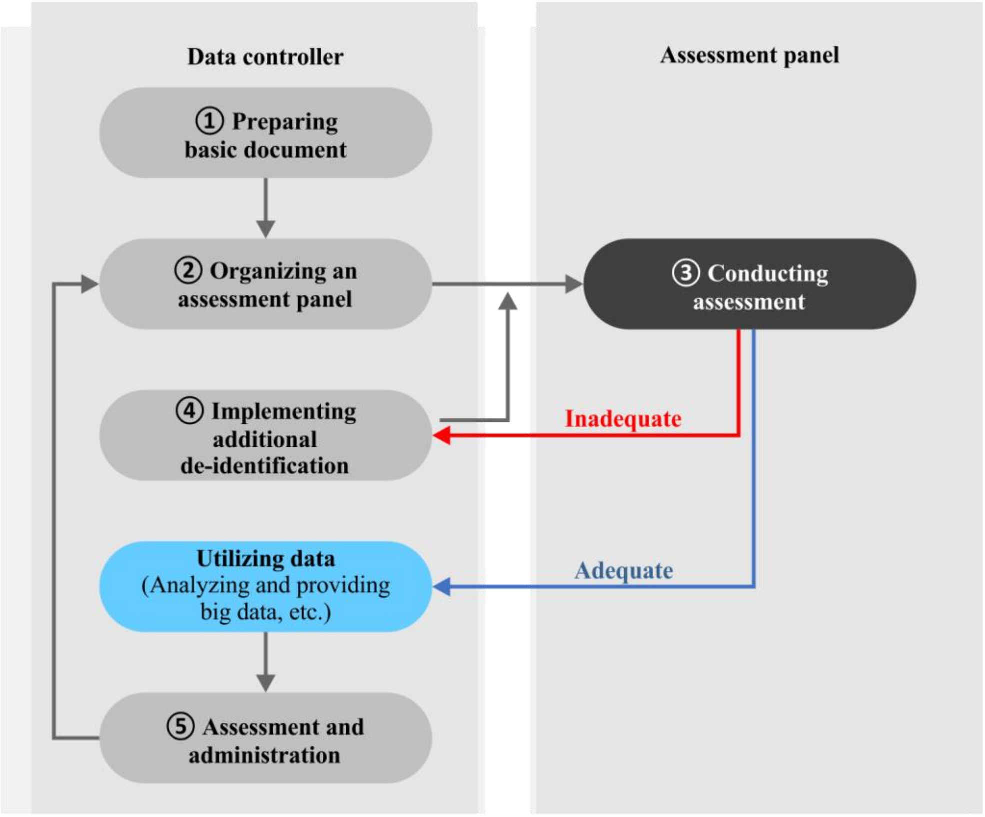
Adequacy assessment procedures

(This annex forms an integral part of this Recommendation.)

This annex provides a model for conducting adequacy assessment procedures [b-KOREA]. See Figure A.1

The following are the steps in the adequacy assessment procedure:

- Preparing basic documents. A data controller shall prepare the basic documents needed for adequacy assessment such as a data statement, de-identification status, and the management level of user organizations. 'user organization' means that an organization intends to utilize de-identified data after the de-identification.
- Organizing an assessment panel. A privacy officer can form the assessment panel or call for a DPO or TTP to perform the assessment.
- Conducting assessment. The assessment panel shall assess the adequacy level of de-identification by utilizing the basic documents that the PII manager prepared.
- Implementing additional de-identification. The data controller shall implement additional de-identification by reflecting opinions of who are participants of the assessment if an assessment result is inadequate.
- Utilizing data. Data can be used or provided for purposes like big data analysis if the assessment of de-identification turns out to be adequate.


Data controller
Assessment panel
① Preparing basic document
② Organizing an assessment panel
③ Conducting assessment
④ Implementing additional de-identification
Inadequate
Adequate
Utilizing data (Analyzing and providing big data, etc.)
⑤ Assessment and administration

X.1148(20)_FA.1

Figure A.1 – Assessment adequacy procedure of de-identification

A.1 Preparing basic documents

A data controller shall prepare basic documents needed for adequacy assessment such as a data statement of an assessment subject, status of de-identification and the level of management of a user organization.

A.2 Organizing an assessment panel

A privacy officer can form the assessment panel. Appoint more than one legal and de-identification expert from an expert pool that is operated by the specialized agencies of each field when commissioning external professionals.

The assessment panel is comprised of members who do not have direct interests in the purpose of data utilization.

A.3 Conducting assessment

The assessment panel assesses the adequacy of de-identification based on basic documents and using the k-anonymity model.

– Preliminary review. Review the basic documents prepared by the data controller and check out through the interview whether personal identification elements are in a dataset, whether the purpose of use and de-identification techniques are appropriate.

– Analysis possibilities of re-identification attempt. Analyze possibilities of re-identification attempts including intention, level of PII protection and capability of the data controller who uses or receives the data.

– Impact analysis of re-identification. Assess the possible impact on a data subject when data is intentionally or unintentionally re-identified.

– Quantitative analysis. Verify the accuracy of a K value provided by a data controller.

– Decide reference value of assessment. The assessment panel comprehensively determines the assessment reference value by taking into consideration possibilities of re-identification, impact of re-identification, results of quantitative analysis and purpose of data use.

– Adequacy assessment. Decide adequacy of de-identification by comparing calculated values resulted from the average reference value and quantitative analysis.

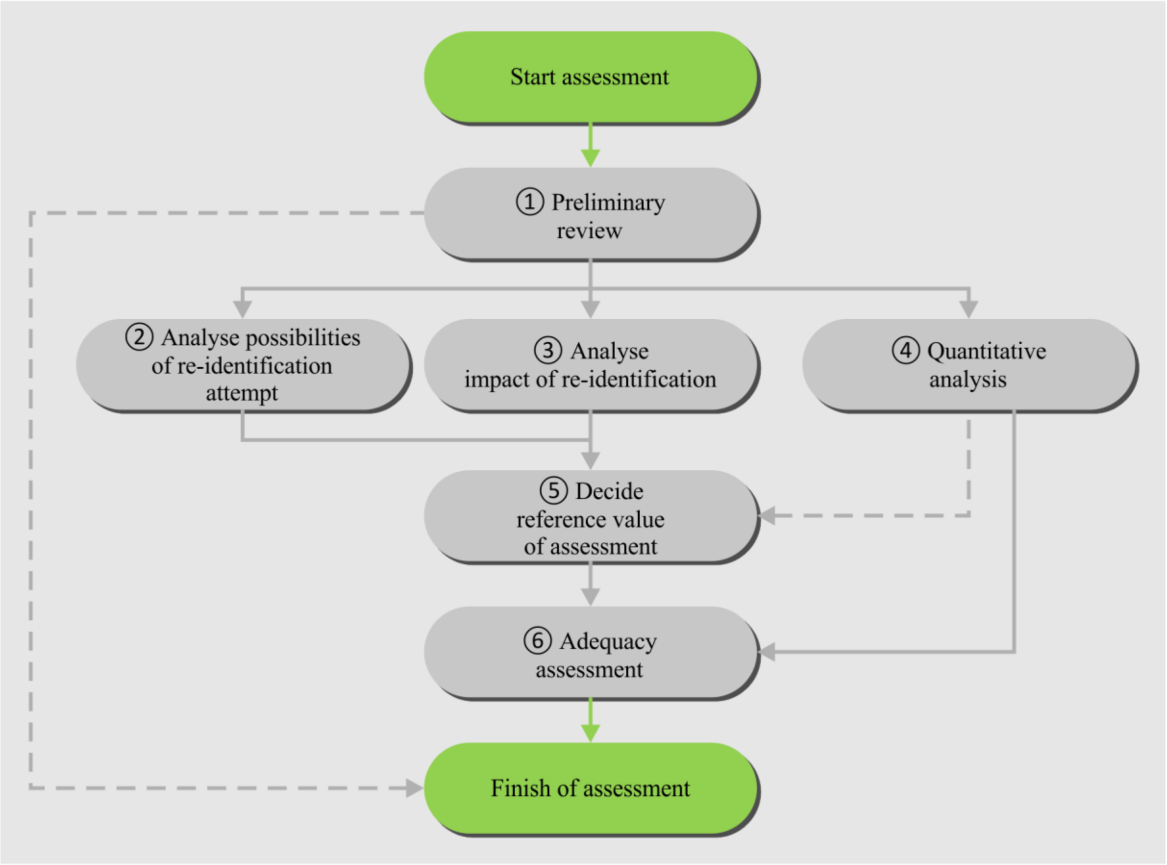

Start assessment
① Preliminary review
② Analyse possibilities of re-identification attempt
③ Analyse impact of re-identification
④ Quantitative analysis
⑤ Decide reference value of assessment
⑥ Adequacy assessment
Finish of assessment
X.1148(20)_FA.2

Figure A.2 – Assessment adequacy procedure

A.4 Additional de-identification measures

- A data controller shall implement additional de-identification measures based on feedback from the assessment panel if the assessment result is inadequate.
- The assessment panel shall proceed to re-assessment once the data controller completes implementation of additional de-identification.

A.5 Data utilization

- Utilize the de-identified data in big data analysis or allow providing it to a third party if de-identification is assessed (reassessed) to be adequate.
- In principle, providing or disclosing data to public or uncontracted data users is prohibited if there is no proper risk mitigation strategy for data release models due to a high risk of re-identification.
- Destroy data once the purpose of using it is fulfilled or it is no longer needed.
- The follow-up management steps should be observed in the process of utilizing data for effective usage in a form of de-identified data.

Annex B

Unstructured de-identification approaches

(This annex forms an integral part of this Recommendation.)

Different from structured data de-identification, de-identification mechanisms for unstructured data are applied to the raw data instead of structured data fields. In the case of the photo below, de-identification means removing faces or replacing them with others such as shown in Figure B.1.

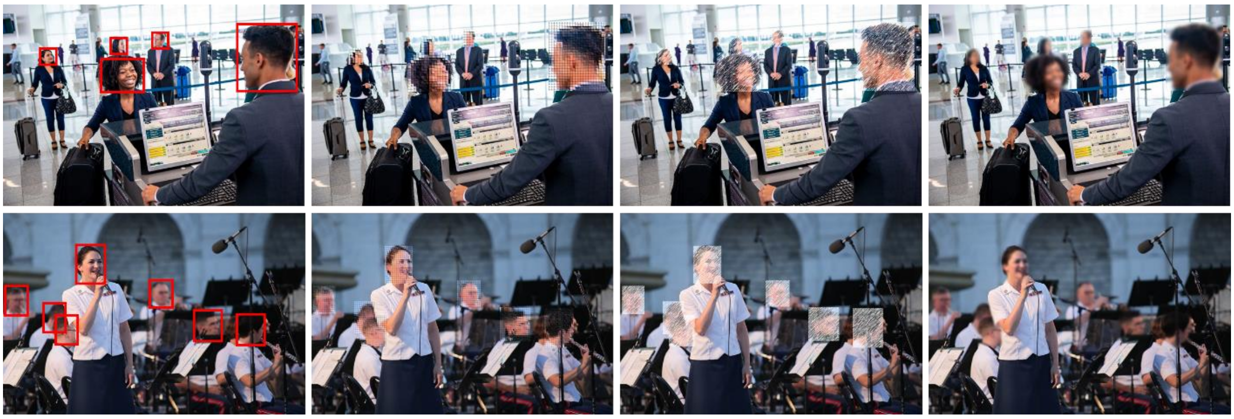


Figure B.11 – A example of face de-identification

There are four types of unstructured data:

1) unstructured text data: web data, report document, blog, news, etc.;

2) unstructured video data: all video data are unstructured, and some tag information provides regularized data;

3) unstructured audio data: all audio data are unstructured, and some tag information or recognized audio is translated into the text data;

4) unstructured log data: machine generated log data are unstructured but usually it has some pattern and can be translated into the structured form.

In order to represent syntactic information for unstructured data including text, voice, image and video, a de-identification processing system should include the following three units:

1) multimedia information detection unit for detecting texted meta information from the input multimedia data:
   - it includes a speech detector that converts voice input into a text to track an object or activity included in the voice input;
   - it includes an optical character recognition detector that extracts characters from an image input;
   - it includes a visual detector that extracts an object or activity included in an input of the image or removing picture from the input of the image or moving picture input;
   - it includes a visual to sentence detector that extracts a text sentence from an image or moving picture input.

2) knowledge-based shaping unit dividing the texted meta information and context information into syntactic representing extrinsic configuration and semantics representing intrinsic information:

- the syntactic information includes source information generating the multimedia data, information of the multimedia data generated by the source, and object detection information extracted from a meaning region;
- the semantic information includes event information included in the meaning region configuring the multimedia data and context information.

3) de-identification unit removes identifiable PII from the knowledge base and texted meta information.

The unstructured data de-identification mechanism should define related requirements and security strength as follows:

- target of de-identification: identify the target object should be protected for application or online services?
- how to carry out de-identification: identify which mechanism should be used for de-identification. What is the level of de-identification (e.g., black box, pixilation, blurring)?
- de-identification vs. re-identification: the need of recovery or re-identification should be determined. When a policy requires an original photo for a crime investigation, can the de-identified photo be recovered?

Appendix I

Examples of typical de-identification techniques

(This appendix does not form an integral part of this Recommendation.)

This appendix provides some examples and descriptions of typical de-identification techniques.

I.1 Statistical tools for de-identification techniques

– Sampling: a process in which a sample of an entire dataset is released, instead of releasing an entire dataset. If a subsample is released, the probability of re-identification can decrease.

– Aggregation: a set of statistical functions that produce the represented value of an entire dataset.

I.2 Cryptographic tools for de-identification techniques

– Deterministic encryption [b-ISO/IEC 11770]: an encryption scheme that always produces the same ciphertext for a given plaintext and key over separate executions of the encryption algorithm.

– Order-preserving encryption [b-AGRAWAL]: an encryption scheme in which numerical ordering of the plaintexts is preserved.

– Homomorphic encryption [b-ISO/IEC 18033-6]: an encryption scheme that allows computations to be carried out on ciphertext, thus generating an encrypted result which matches the result of operations to be performed on the plaintext, when decrypted.

– Format-preserving encryption [b-NIST 800-38G]: an encrypting scheme in which the ciphertext is in the same format as the plaintext.

– Homomorphic secret sharing [b-ISO/IEC 18033-6]: a type of secret sharing algorithm in which the secret is encrypted using a homomorphic encryption.

I.3 Suppression techniques

– Masking: the process of replacing a field with a value or removing it. The example of suppression technique includes replacing a phone number with asterisks or a randomly generated pseudonym.

– Local suppression: a process that suppresses or removes specific values of attributes from selected records. Removing the data increases privacy protection but may decrease the utility of the dataset.

– Record suppression: a process that involves removing an entire record or records from a dataset.

I.4 Pseudonymization techniques

A process that both removes the association with a data subject and adds an association between a particular set of characteristics relating to the data subject and one or more pseudonyms. Typically, pseudonymization is implemented by replacing direct identifiers with a pseudonym, such as a randomly generated value. Examples of direct identifiers include names, email addresses, and government issued numbers. All direct identifiers and potentially additional or all remaining identifying attributes are replaced with a pseudonym.

I.5 Generalization techniques

– Rounding: a process of replacing a numerical value by another value that is approximately equal but has a shorter, simpler or more explicit representation.

- Top and bottom coding: a process for which the attribute whose values are above an upper bound (or lower bound) are set as a threshold on the largest (or smallest) value possible.

I.6 Randomization techniques

- Noise addition: a process in which a random value that cannot be predicted is added to a selected attribute of a dataset.
- Permutation: a process for exchanging the values of a selected attribute across the records in a dataset without modification.
- Micro-aggregation: a process in which all values of continuous attributes are replaced with their averages computed in a certain algorithmic way.

I.7 Synthetic data

Synthetic data is an approach that artificially generates micro data to represent a predefined statistical data model. By definition, synthetic datasets do not contain data collected from existing data subjects, but they look real for their intended purpose.

Appendix II

De-identification process approaches

(This appendix does not form an integral part of this Recommendation.)

This appendix provides some examples and details of de-identification process approaches.

II.1 Data-centric de-identification approach

Given that de-identification techniques modify the original data to prevent disclosure of PII, tension clearly arises between utility and privacy. The challenge is to protect privacy with minimum loss of accuracy: ideally, data users should run their analyses on the de-identified data without losing accuracy with respect to the results of those analyses when run on the original data.

Perfect de-identification is difficult in practice without compromising the utility of the dataset. In big data, this problem increases due to the amount and variety of data. On the one hand, low-level de-identification (e.g., de-identification only suppressing direct identifiers) is usually not enough to ensure non-identifiability. On the other hand, too strong de-identification may prevent linking data on the same individual (or on similar individuals) that come from different sources and, thus, thwart many of the potential benefits of big data.

This clause describes two approaches for data-focused de-identification to deal with the tension between utility and privacy. Data-use specific and generic utility measures can be used to address how to measure the utility of a de-identified released dataset.

II.1.1 Utility-first de-identification approach

In big data, information about an individual is often gathered from several independent sources. Hence, the ability to link records that belong to the same (or of the same type/similar) individual is central in big data creation.

In the utility-first de-identification approach, a de-identification technique with a heuristic parameter choice and with suitable utility preservation properties is run on the micro-dataset and, after that, the risk of disclosure is measured. Therefore, the utility-first de-identification approach is slow and lacks formal privacy guarantees. For instance, the risk of re-identification can be estimated empirically by attempting record linkage between the original and the de-identified datasets. If the extant risk is deemed too high, the de-identification technique must be re-run with more privacy-stringent parameters and probably with more utility sacrifice, iteratively changing parameters until empirical disclosure risk is low enough, as usual in official statistics.

Of course, while linkability is desirable from the utility point of view, it is also a privacy threat: the accuracy of linkages should be significantly less in de-identified datasets than in original datasets. The amount of linkability compatible with a de-identification technique or with a de-identification privacy model determines whether and how an analyst can link independently de-identified data (under that technique/model) that correspond to the same individual.

II.1.2 Privacy-first de-identification approach

A privacy model is enforced with a parameter that guarantees an upper bound on the re-identification disclosure risk and perhaps also on the attribute disclosure risk. Model enforcement is achieved by using a model-specific de-identification technique with parameters that derive from the model parameters. Well-known privacy models include k-anonymity and its extensions, as well as ε-differential privacy, that often leads to poor data utility/linkability.

In the privacy-first de-identification approach, if the utility of the resulting de-identified data is too low, then the privacy model in use should either be enforced with an alternative de-identification technique that is less utility-damaging, or a less strict privacy parameter should be chosen, or even a different de-identification privacy model should be resorted to.

II.2 Role-centric de-identification approach

This clause describes three types of approaches which perform each other's roles and responsibilities in the de-identification process. Role-focused approach can be broadly characterized as answering 'who', 'what', and 'where and how' questions:

– Who has access to the data?

– What analyses may or may not be conducted?

– Where is the data access/analysis to be carried out and how is access obtained?

II.2.1 Centralized de-identification

The process on statistical disclosure control focuses on centralized de-identification, performed by a data controller who has access to the entire original dataset. This centralized approach has some advantages and downsides as shown in Table II.1.

Table II.1 – Characteristics of centralized de-identification

	Details
Advantages	• Individuals do not need to de-identify the data records they provide. The data controller, who has more computational resources and probably more expertise in de-identification, can be expected to adequately de-identify the entire dataset. • The data controller has a global view of the original dataset and, thus, is in the best position to optimize the trade-off between data utility and extant disclosure risk.
Disadvantages	• The data controller must be trusted by all parties providing original data (because the controller has access to all original data). While this is not a problem in official statistics, where the data controller is a national statistical institute, it can be a major hurdle in a typical big data scenario, for instance when the data controller assembling several data sources is merely a private company (e.g., a data broker). • Especially in the case of big data, de-identification can be too heavy a computational burden for a single controller. • Many controllers are involved in a single big data processing scenario, thus, making the centralised approach unmanageable.

The local de-identification approaches and collaborative de-identification is a complement to the above advantages and downsides.

II.2.2 Local de-identification

Local de-identification is an alternative disclosure limitation approach suitable for scenarios (including big data) where individuals (data subjects) do not trust (or trust only partially) the data controller assembling the data. Each subject de-identifies its own data before handing them to the data controller.

With privacy protection in mind, the data collected by a given source should be de-identified at the source before being made available. However, independent de-identification by each source has more information loss than in centralized de-identification because subjects de-identify their data without seeing the data of other subjects. That is, the subjects lack a global view of the dataset, which makes it difficult for them to find a good trade-off between the disclosure risk limitation achieved and the information loss incurred.

II.2.3 Collaborative de-identification

The collaborative de-identification process combines the low utility loss of centralized de-identification and the high subject privacy of local de-identification. A problem with centralized de-identification is that, if a data subject does not trust the data controller to properly use and/or de-identify his/her data, he/she may decide to provide false data (hence causing a response bias) or no data at all (hence causing a non-response bias). Therefore, subjects could collaborate to determine the disclosure risk associated to their data and then locally apply the right level of protection in a distributed and collaborative manner, which seeks two main properties:

- It incurs no more information loss than the dataset that would be obtained with the centralized approach for the same privacy level. It outperforms the local approach in that it yields less information loss.
- Neither the data subjects nor the data controller gain more knowledge about the confidential attributes of any other specific data subject than the knowledge contained in the final de-identified dataset. It outperforms the centralized approach by also offering privacy versus the data collector.

In addition, the collaborative approach could lead to protocols that work smoothly without external enforcing mechanisms. In micro-data de-identification, the privacy protection obtained by a subject affects the privacy protection that others get. For improving co-utility in collaborative approach, secure multiparty transformation is required to electronic protocols that enable two or more parties to carry out a transformation that involves both of their datasets in such a way that no party needs to explicitly hand a dataset to any of the others. Because secure multiparty transformation allows for queries to be transformed without the need for all data storage to be centralized, it diminishes the harm from data breach and allows computations across parties that do not fully trust each other. Multiparty computations can offer both better privacy and utility in certain contexts.

Bibliography

[b-ISO/IEC 11770] ISO/IEC 11770 (all parts), *Information technology – Security techniques – Key management*.

[b-ISO/IEC 18033-6] ISO/IEC 18033-6, *Information technology security techniques – Encryption algorithms – Part 6: Homomorphic encryption.*

[b-ISO/IEC 20889] ISO/IEC 20889 (2018), *Privacy enhancing data de-identification terminology and classification of techniques.*

[b-ISO/IEC 27001] ISO/IEC 27001 (2018), *Information technology – Security technique – Information security management systems.*

[b-ISO/IEC 29100] ISO/IEC 29100 (2011), *Information technology – Security technique – Privacy framework.*

[b-NIST 800-38G] NIST Special Publication 800-38G (2016), *Recommendation for Block Cipher Modes of Operation: Methods for Format-Preserving Encryption.*

[b-NISTIR 8053] NISTIR 8053 (2015), *De-Identification of Personal Information.*

[b-AGRAWAL] Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. (2004), *Order preserving encryption for numeric data, SIGMOD '04 Proceedings of the 2004 ACM SIGMOD international conference on Management of data, Paris, France, June, pp 563-574.*

[b-KOREA] Korean Ministry of the Interior, *Guidelines on De-identification Measures, June 2016.*
<http://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000821178&fileSn=2&nttId=7187&toolVer=&toolCntKey_>
Last accessed 26 July 2019)
<https://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000827161&fileSn=0>
(English, last accessed 12 December 2020)

[b-UKAN] UK Anonymization Network, *The anonymisation decision-making framework, 2016*
<https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	**Data networks, open system communications and security**
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2020