ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T X.1771 (04/2024)

SERIES X: Data networks, open system communications and security

Data security – Data protection (II)

Security guidelines for combining de-identified data using trusted third party

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	X.1000-X.1099
SECURE APPLICATIONS AND SERVICES (I)	X.1100-X.1199
CYBERSPACE SECURITY	X.1200-X.1299
SECURE APPLICATIONS AND SERVICES (II)	X.1300-X.1499
CYBERSECURITY INFORMATION EXCHANGE	X.1500-X.1599
CLOUD COMPUTING SECURITY	X.1600-X.1699
QUANTUM COMMUNICATION	X.1700-X.1729
DATA SECURITY	X.1750-X.1799
Big Data Security	X.1750-X.1759
Data protection (II)	**X.1770-X.1799**
INTERNATIONAL MOBILE TELECOMMUNICATIONS (IMT) SECURITY	X.1800-X.1839
METAVERSE AND DIGITAL TWIN SECURITY	X.2000-X.2199
SOFTWARE SUPPLY CHAIN SECURITY	X.2150-X.2199
ARTIFICIAL INTELLIGENCE (AI) / MACHINE LEARNING (ML) SECURITY	X.2200-X.2249

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

Recommendation ITU-T X.1771

Security guidelines for combining de-identified data using trusted third party

Summary

De-identification is a process to remove the association between a set of identifiable attributes and the data principal. Appropriate de-identification techniques should be used by an organization to balance between its desires to protect data as the data controller and to use the data in new and innovative ways.

Recommendation ITU-T X.1771 provides security guidelines for information and communication technology (ICT) service providers acting as data controllers and for trusted third parties fulfilling the role of combining de-identified datasets. The purpose of these security guidelines is to reduce risks, protect data and promote the utility of the dataset.

History [*]

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T X.1771	2024-04-29	17	11.1002/1000/15886

Keywords

Combining de-identified datasets, combining key, controls, de-identification, de-identified dataset, model for combining de-identified datasets, procedures for combining de-identified datasets.

[*] To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

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/software copyrights, 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 appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T X.1771

Security guidelines for combining de-identified data using trusted third party

1 Scope

De-identification can serve as a solution for an organization to balance between its desires to protect data and to harness the data for innovative purposes. The de-identification techniques described in [ITU-T X.1148] can be used to produce a de-identified dataset from a dataset having an identifying attribute. In addition, there is a need for combined de-identified datasets from different organizations, which can be used for example for statistical calculation, scientific analysis, and other purposes.

This Recommendation provides security guidelines for securely combining de-identified datasets from different organizations. It describes use cases and two models for combining de-identified datasets, identifies related entities, and specifies procedures to combine de-identified datasets submitted from different organizations acting as data controllers or data processors. In addition, this Recommendation proposes organizational and technical controls for trusted third parties which are responsible for combining de-identified datasets for information and communication technology (ICT) service providers.

This Recommendation does not address issues related to regulation.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T X.1148] Recommendation ITU-T X.1148 (2020), *Framework of de-identification process for telecommunication service providers*.

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

3.1.3 control [b-ISO/IEC 27000]: A measure that is modifying risk.

NOTE 1 – Controls include any process, policy, device, practice, or other actions which modify risk.

NOTE 2 – It is possible that controls do not always exert the intended or assumed modifying effect.

3.1.4 data controller [ITU-T X.1148]: Stakeholder (or privacy stakeholder) that determines the purposes and means for processing data other than natural persons who use data for personal purposes.

3.1.5 data processor [ITU-T X.1148]: Stakeholder that processes data on behalf of and in accordance with the instructions of a data controller.

3.1.6 data principal [b-ISO/IEC 20889]: Entity to which data relates.

NOTE – The term "data principal" is broader than "PII principal" (or "data subject" as used elsewhere), and is able to denote any entity such as a person, an organization, a device, or a software application.

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

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

3.1.9 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.

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

3.1.11 direct identifier [b-ISO/IEC 20889]: Attribute that alone enables unique identification of a data principal within a specific operational context.

NOTE – Here and throughout, the operational context includes the information that the entity processing (e.g., de-identifying) the data possesses, together with information that third parties and potential attackers can possess or that is in the public domain.

3.1.12 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 of [b-ISO/IEC 20889] for a discussion of how this definition relates to those given in other standards.

3.1.13 indirect identifier [b-ISO/IEC 20889]: Attribute that, together with other attributes that can be in the dataset or external to it, enables unique identification of a data principal within a specific operational context.

3.1.14 personally identifiable information [b-ISO/IEC 29100]: Any information that (a) can be used to identify the personally identifiable information (PII) principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal.

NOTE – To determine whether a PII principal is identifiable, account should be taken of all the means which can reasonably be used by the privacy stakeholder holding the data, or by any other party, to identify that natural person.

3.1.15 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.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 combined de-identified dataset: Dataset resulting from the implementation of a combining process with the input of de-identified datasets submitted from different organizations.

3.2.2 combining organization: A trusted third party responsible for combining de-identified datasets submitted from different organizations.

3.2.3 combining process: Process of combining de-identified datasets from organizations to produce extensive data in the characteristic of volume that requires a scalable technology for efficient storage, manipulation, and analysis, and to remove the association between a set of identifying attributes and the data principal.

3.2.4 combining key: Secret information shared by relevant organizations for combining de-identified datasets.

NOTE – The combining key is derived from a hashed outcome of secret information previously shared among relevant stakeholders to uniquely identify information belonging to a data principal using one-way hash function from among the datasets received from applicant organizations.

3.2.5 **data record**: Set of attributes concerning a single data subject.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ICT	Information and Communication Technology
PII	Personally Identifiable Information
TTP	Trusted third party

5 Conventions

This Recommendation uses the following conventions:

The keywords "**is required to**" or "**shall**" indicate a requirement which must be strictly followed and from which no deviation is permitted, if conformance to this Recommendation is to be claimed. This keyword is interchangeable with "shall" in this Recommendation.

The keywords "**should**" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance. This keyword is interchangeable with "should" in this Recommendation.

6 Overview

Major benefits can be obtained from processing data, including so-called "big data". However, where this data includes personally identifiable information (PII) data, processing this data needs to consider data protection as described in [b-ISO/IEC 29100]. The appropriate use of de-identification techniques is important to enable the exploitation of the benefits of data processing while maintaining compliance with the relevant privacy principles presented in [b-ISO/IEC 29100].

It is well known that de-identification can be used to provide a balance between protecting data collected from a data subject and an organization's desire to use data in new and innovative ways. The appropriate use of de-identification techniques should be employed to support this balance.

The purpose of the de-identification process is to remove the association between a set of identifying attributes and the data principal. A de-identified dataset is a dataset resulting from the application of a de-identification process where a dataset is a collection of data that does not contain any PII data.

A de-identification technique is a method for transforming a dataset with the objective of reducing the extent to which data could be associated with individual data principals.

The identifier is a set of attributes in a dataset that enables unique identification of a data principal within a specific operational context. The identifier consists of direct identifiers and indirect identifiers. All values that are not direct identifiers are indirect identifiers. Examples of direct identifiers include an individual's name (including the name of a substitute decision maker, next of kin, etc.), telephone number, facsimile number, home address, electronic mail address, health insurance number, social insurance number and medical record number. Examples of indirect identifier are an individual's age or date of birth, race, salary, educational attainment, occupation, marital status and zip code.

To use the de-identified data in innovative ways, there is a need to combine de-identified data provided from different organizations into a combined de-identified dataset.

6.1 Combining de-identified datasets

Figure 1 shows an overview of the de-identification process and a basic use case for producing de-identified datasets in [ITU-T X.1148].

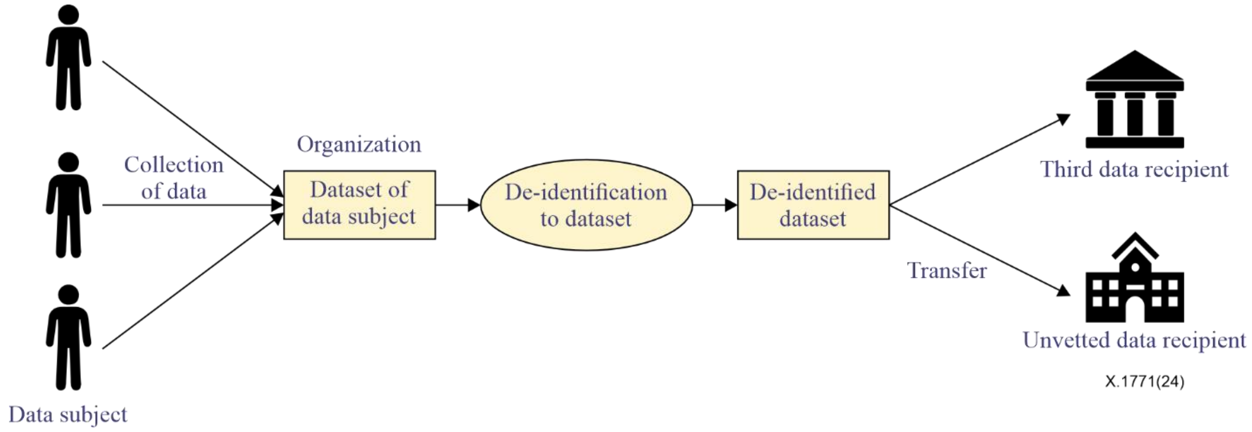

Figure 1 – Overview of the de-identification process and basic use case for producing de-identified datasets

Data are collected from data subjects, the persons to whom the data relate. These data are transformed into a dataset containing PII data. De-identification creates a new dataset, also known as de-identified dataset, considered to have no PII data. The de-identification techniques described in [ITU-T X.1148] can be used to produce the de-identified dataset in an organization.

This de-identified dataset may be internally used by an organization instead of the original dataset to decrease data protection risks. The de-identified dataset may also be provided to trusted data recipients who are bound by additional administrative controls such as data use agreements. Alternatively, the de-identified data might be made broadly available to a larger (potentially limitless) number of unknown and unvetted data recipients, for example, by publishing the de-identified data on the Internet.

Figure 2 illustrates the extended use case for combining de-identified datasets from different organizations.

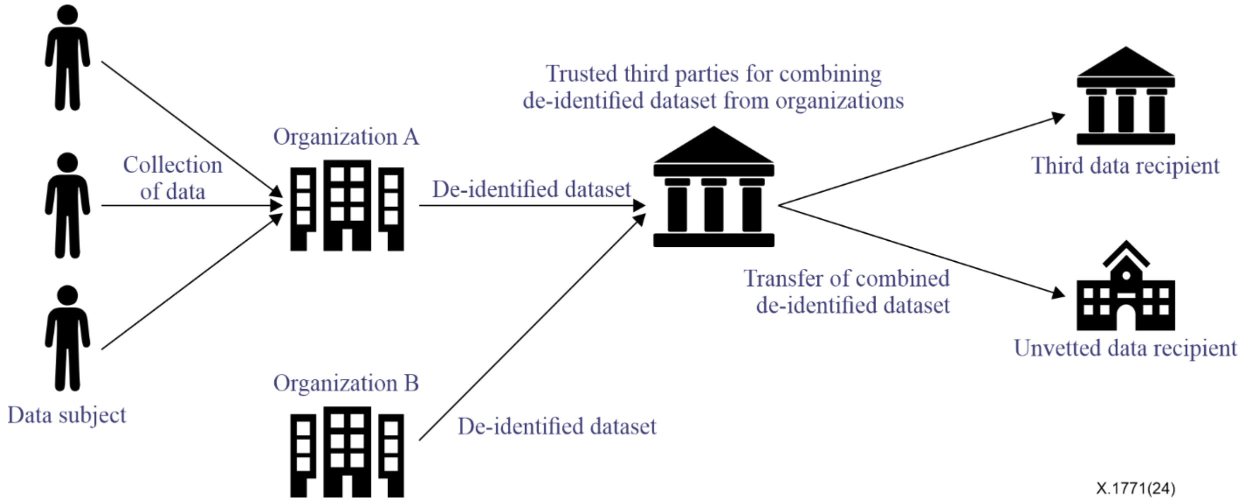

Figure 2 – Extended use case for combining de-identified datasets submitted from different organizations

The de-identified dataset of an organization is transferred to a trusted third-party combining organization. The role of the combining organization is to combine de-identified datasets from different organizations to produce a combined de-identified dataset. To decrease data protection risks, the combining function for de-identified datasets from different organizations is conducted by the combining organization instead of by each organization.

6.2 Generation of a combining key

A combining key is the information which can be used to identify the de-identified datasets in the two applicant organizations, which belong to a same data subject.

Each combining applicant which provides the de-identified dataset to a combining organization negotiates with a combining key generation organization on matters related to the creation of a combining key and receives secret information (i.e., a salt value) necessary for generating a combining key from the combining key generation organization.

When generating a combining key, the input items for creating the combining key, encoding method, and key generation algorithm (i.e., hash algorithm) between the applicant organizations shall be the same.

In general, the combining key is used to identify a specific data subject, whose attribute has the same name, phone number, and date of birth. The combining key generation algorithm uses a one-way hash algorithm so that a specific data subject cannot be identified as a combining key generation item.

A typical example of generating a combining key is shown in Figure 3. The hash algorithm can be used to generate the combining key. The input value of the hash algorithm comprises an identifying attribute of the same data subject such as name, unique identifier, birth date of the data subject, the 'salt' shared between the applicant organizations and the combining key generation organization.

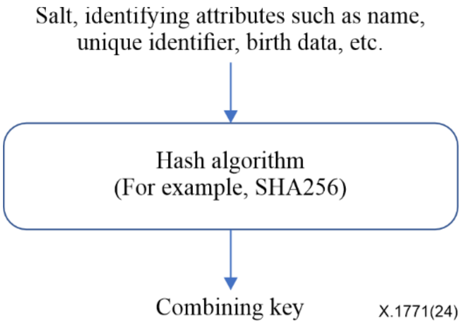
Salt, identifying attributes such as name,
unique identifier, birth data, etc.
Hash algorithm
(For example, SHA256)
Combining key
X.1771(24)

Figure 3 – Typical example of generation of a combining key

7 Use cases for combining de-identified datasets from organizations

Data about data subject A can be used for the purpose that an organization informs data subject A about when collecting data from data subject A. Once the data is de-identified, that data can be used without separate consent required from the data subject for other purposes such as for example, statistical purposes, scientific research purposes, or archiving purposes in the public interest [b-PIPC]. If data is de-identified, that de-identified data can be transferred to the third party without further consent required from data subject A according to regulations.

Figure 4 describes a use case for combining de-identified datasets from a national cancer hospital, a national health insurance agency, and a national statistics agency. In this case, de-identified clinical data of lung cancer patients at the National Cancer agency, de-identified medical data of data subject at the National Health Insurance agency and de-identified death data at the National Statistical Office were combined by the single trusted third party. A large, combined dataset is analysed to produce useful public health information. The results of analysis of combined de-identified datasets can produce information on trends and causes of death of lung cancer patients. For example, here it was shown that there was a 38% chance that lung cancer patients would die within one year.

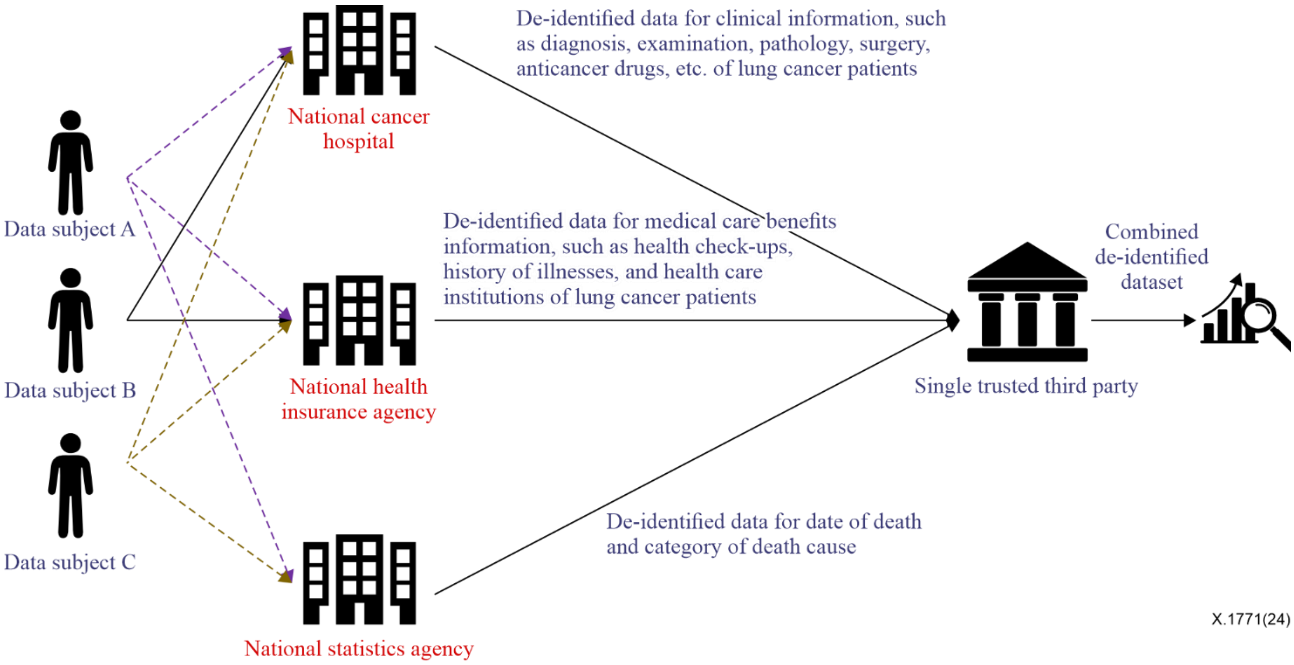
De-identified data for clinical information, such as diagnosis, examination, pathology, surgery, anticancer drugs, etc. of lung cancer patients
National cancer hospital
Data subject A
De-identified data for medical care benefits information, such as health check-ups, history of illnesses, and health care institutions of lung cancer patients
Combined de-identified dataset
Data subject B
National health insurance agency
Single trusted third party
Data subject C
De-identified data for date of death and category of death cause
National statistics agency
X.1771(24)

Figure 4 – A use case for combining de-identified datasets from three organizations

8 Models for combining de-identified datasets

There are two types of models for combining de-identified datasets from organizations: model A with a single trusted third party and model B with two trusted third parties.

8.1 Model with one trusted third party: combining organization

Figure 5 illustrates model A with a single trusted third party, a combining organization, for combining de-identified datasets from organizations. In this model, it is assumed that a data subject uses a service provided by an Organization A and an Organization B. Data about the data subject can be collected by both organizations each as a data controller for the specific purpose. Organization A uses the de-identification process presented in [ITU-T X.1148] to obtain a de-identified dataset for some or all data subjects. Organization B also uses a similar process. The de-identified dataset contains de-identified data about the data subject. Both Organization A and Organization B can transfer the de-identified dataset without requiring the consent of data subject. In this case, a single trusted third party as the combining organization can combine the de-identified datasets from applicant Organization A and Organization B to obtain a big data record about each data subject. The combined data de-identified datasets can be analysed to derive a useful statistical data.

In order for the combining organization to identify the data that belongs to the data subject from the de-identified dataset, the combining key that can be derived from data attributes about the data subject should be transferred to link a data record about the data subject from Organization A with that from Organization B. The combining key can be derived as a hash value of a secret value provided by a single trusted third party, or a unique identifier about the data subject such as a national identifier, or an e-mail address, etc.

The combining organization uses the combining key to combine the data record from Organization A with that from Organization B to obtain the big data record about the data subject.

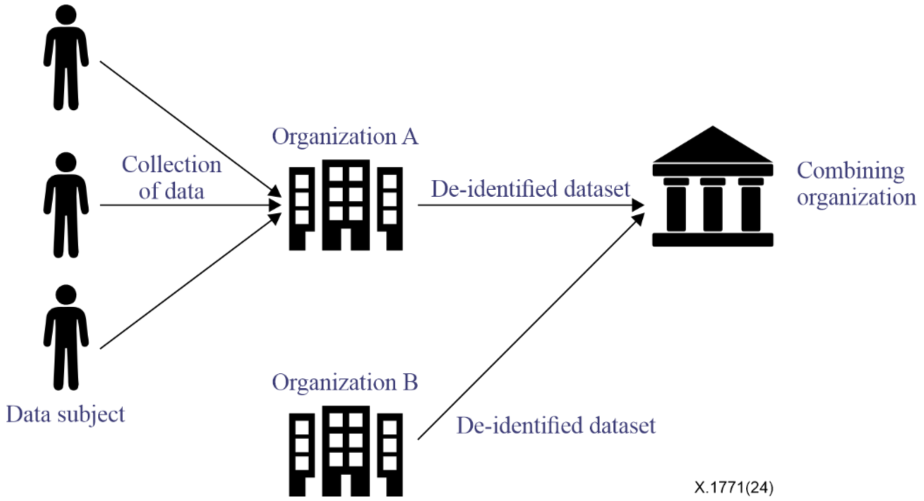

Figure 5 – Model A with a single trusted third party

8.2 Model with two trusted third parties: key management authority and combining organization

Figure 6 illustrates model B with two trusted third parties: a key management authority and a combining organization, for combining de-identified datasets from different organizations.

In this model, it is assumed that a data subject uses services provided by an Organization A and an Organization B. The combining organization combines two datasets from Organization A and Organization B to obtain the big data record for the data subject. The key management authority provides a secret value to each organization to obtain the combining key. The key management agency uses this combining key to provide the combining key related information to the combining organization that enables combining of the data from de-identified datasets.

Data about the data subject can be collected by both Organization A and Organization B as a data controller for a specific purpose. Organization A uses the de-identification process presented in [ITU-T X.1148] to obtain a de-identified dataset for some or all data subjects. Organization B also uses similar process. The de-identified dataset contains de-identified data about the data subject. Both Organization A and Organization B can share the de-identified dataset without consent of the data subject. In this case, the combining organization as a trusted third party can combine the de-identified datasets from applicant Organization A and Organization B to obtain the big data records about the data subjects. In order for the combining organization to identify the data that belong to the data subject from the de-identified datasets, the combining key related information that can be calculated by the key management authority and Organization A and Organization B, which can be derived from a data attribute about the data subject should be transferred to allow the combining organization to enable linking of the data record about the data subject from Organization A with that from Organization B.

The combining key related information that can be derived from the combining key is a hash value of a secret value provided by the key management authority, a unique identifier about the data subject such as a national identifier, an e-mail address, etc. The combining organization uses the combining key related information to combine the data record from applicant Organization A with that from Organization B to obtain the big data record about the data subject.

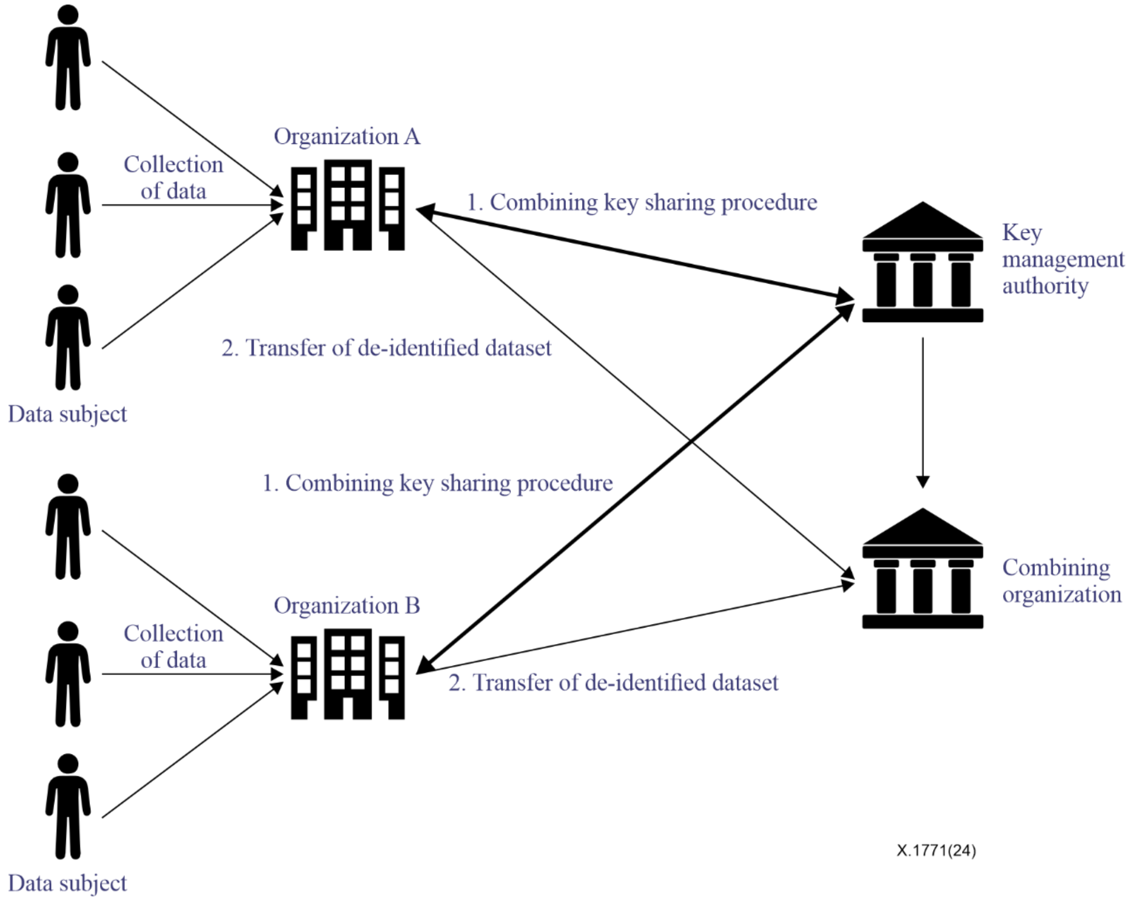

Figure 6 – Model B with two trusted third parties

9 Procedures to combine de-identified datasets

Overall, the procedure for combining data from multiple data controllers consists of 4 stages: preparing the application for combining data, creating the combining key and transferring data to be combined, additional processing and requesting for export, and export and follow-up management.

1) During the stage for **preparing the application for combining data**, an organization prepares for the combination process through de-identification or by preparing the relevant applications, which must be submitted to the combining organization.

2) During the stage for **creating combining key and transferring data to be combined**, an organization discusses and agrees with the combining organization, among others, on the combination schedule and the method of transfer. In addition, the organization creates a combination key using the relevant information provided by the combining organization for managing the combination keys.

3) During the stage for **combining the de-identified datasets and requesting for export**, the organization shall conduct additional de-identification on the combined de-identified data in a designated location on the combining organization's premises before exporting such data. In order to request for export of the combined de-identified dataset, the combining organization's Data Transfer Evaluation Committee shall first assess the validity of export of the combined de-identified dataset.

4) During the stage for **export and follow-up management**, the organization may process the exported dataset from the third stage for the original purpose of applying for the combination. In addition, the organization shall comply with the safe handling requirements applicable to the processing of the de-identified dataset.

9.1 Procedures for the model with one trusted third party

Figure 7 describes the procedure for combining de-identified datasets using the model with one trusted third party.

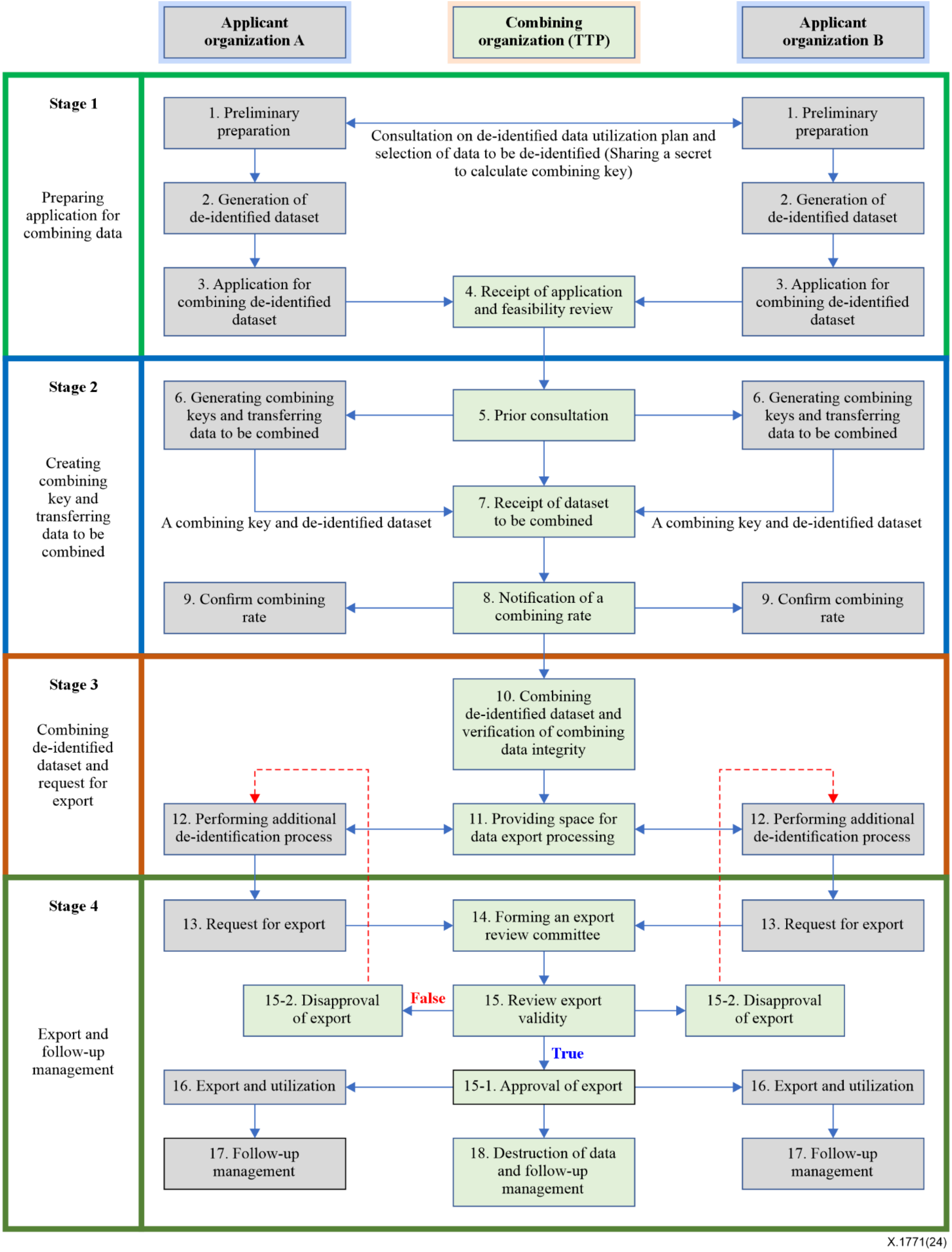
Applicant organization A
Combining organization (TTP)
Applicant organization B
Stage 1
Preparing application for combining data
1. Preliminary preparation
Consultation on de-identified data utilization plan and selection of data to be de-identified (Sharing a secret to calculate combining key)
1. Preliminary preparation
2. Generation of de-identified dataset
2. Generation of de-identified dataset
3. Application for combining de-identified dataset
4. Receipt of application and feasibility review
3. Application for combining de-identified dataset
Stage 2
Creating combining key and transferring data to be combined
6. Generating combining keys and transferring data to be combined
5. Prior consultation
6. Generating combining keys and transferring data to be combined
A combining key and de-identified dataset
7. Receipt of dataset to be combined
A combining key and de-identified dataset
9. Confirm combining rate
8. Notification of a combining rate
9. Confirm combining rate
Stage 3
Combining de-identified dataset and request for export
10. Combining de-identified dataset and verification of combining data integrity
12. Performing additional de-identification process
11. Providing space for data export processing
12. Performing additional de-identification process
Stage 4
Export and follow-up management
13. Request for export
14. Forming an export review committee
13. Request for export
15-2. Disapproval of export
False
15. Review export validity
15-2. Disapproval of export
True
16. Export and utilization
15-1. Approval of export
16. Export and utilization
17. Follow-up management
18. Destruction of data and follow-up management
17. Follow-up management
X.1771(24)

Figure 7 – Procedure for combining de-identified datasets using the model with one trusted third party

In stage 1 (Preparing the application for combining datasets), there are 4 steps:

Step 1 (Preliminary preparation) – Two organizations A and B conduct preliminary preparations for the combination of de-identified datasets through mutual consultation where selection of de-identified data suitable for the purpose of combination, sharing of the secret known as 'salt' to generate a combination key, the level of de-identification that is considered necessary to achieve, the purpose, etc. are agreed. They decide which items are to be used when generating a combination key among the datasets they have in common. They also share the secret known as 'salt' to calculate the combining key.

Step 2 (Generation of de-identified datasets) – The two organizations create a series of non-overlapping numbers (so called serial numbers) for each data record that is subject to be combined. They conduct de-identification processing for the remaining data except for the item for the combination key.

Step 3 (Application for combining de-identified datasets) – The two organizations fill out the application for combination of de-identified datasets, prepare related documents, and submit the application for combination to the combining organization.

Step 4 (Receipt of application and feasibility review) – After receiving the applications from two organizations as a combining applicant and reviewing the feasibility of the combination, a combining organization checks the validity of the application and attached documents, and if insufficient, it also requests the applicant organizations to submit the missing documents.

The applicant organization should submit additional documents, etc. which are requested by the combining organization. If the combining organization recommends the level of de-identification processing, this should be reflected unless there are special rationales.

In stage 2 (Creating combining key and transferring datasets to be combined), there are 5 steps:

Step 5 (Prior consultation) A combining organization consults in advance with the applicant organizations on matters related to combining the de-identified datasets and a combining key generation.

Step 6 (Generation of combining keys and transferring datasets to be combined) The applicant organization generates the combining key by applying the items agreed upon in the preliminary preparation stage.

Step 7 (Receipt of dataset to be combined) The combining organization receives the information transmitted by the applicant organization.

Step 8 (Notification of a combining rate) If the combining applicant organization requests a prior combining rate in the consultation stage with the combining key generation organization, the combining organization notifies the combining rate to the applicant organization.

Step 9 (Confirmation of combining rate) The applicant organization may suspend the application for combining de-identified datasets, if necessary, after confirming the combining rate.

In stage 3 (Combining de-identified datasets and request for export), there are 3 steps:

Step 10 (Combining de-identified datasets and verification of combined dataset integrity) The combining organization performs the combining function using the de-identified datasets provided by the applicant organization. After combining, the combined dataset is verified by checking the integrity of the combined de-identified dataset.

Step 11 (Providing space for dataset export processing) The combining organization provides the export processing space so that the combining applicant organization can perform additional de-identification to export the combined de-identified dataset.

Step 12 (Performing additional de-identified process) The applicant (organization) processes the combined information under an additional pseudonym or with anonymity for export in the export processing space installed within the specialized association, and requests necessary support, such as advice, from the organization specializing in association.

In stage 4 (Export and follow-up management), there are 6 steps:

Step 13 (Request for export) The applicant organization should perform additional de-identification processes, if needed, in the export processing space installed within the combining organization, and request necessary support, such as advice, from the combining organization.

Step 14 (Forming an export review committee) When a combining applicant organization requests for export, the combining organization forms an export review committee to examine whether or not it is appropriate to export the combined de-identified dataset.

Step 15 (Review export validity and approval or disapproval of export) If the combining applicant organization determines that it is necessary for export review or if there is a request from the export review committee, the combining applicant organization shall submit additional datasets or attend and explain the relevant details to the export review committee.

Step 16 (Export and utilization) The applicant organization can use the exported combined de-identified dataset for the purpose of the combination.

Step 17 (Follow-up management) The combining applicant organizations shall monitor and conduct follow-up management.

Step 18 (Destruction of relevant datasets and follow-up management) When the combining process is completed, the relevant history is managed and immediately deleted (combining key, etc.) except for information that is required to be separate by regulation. Follow-up management and supervision by the applicant organization is carried out, and the applicant organization shall respond to the dataset requested by the combining organization for follow-up management and supervision.

9.2 Procedures for the model with two trusted third parties

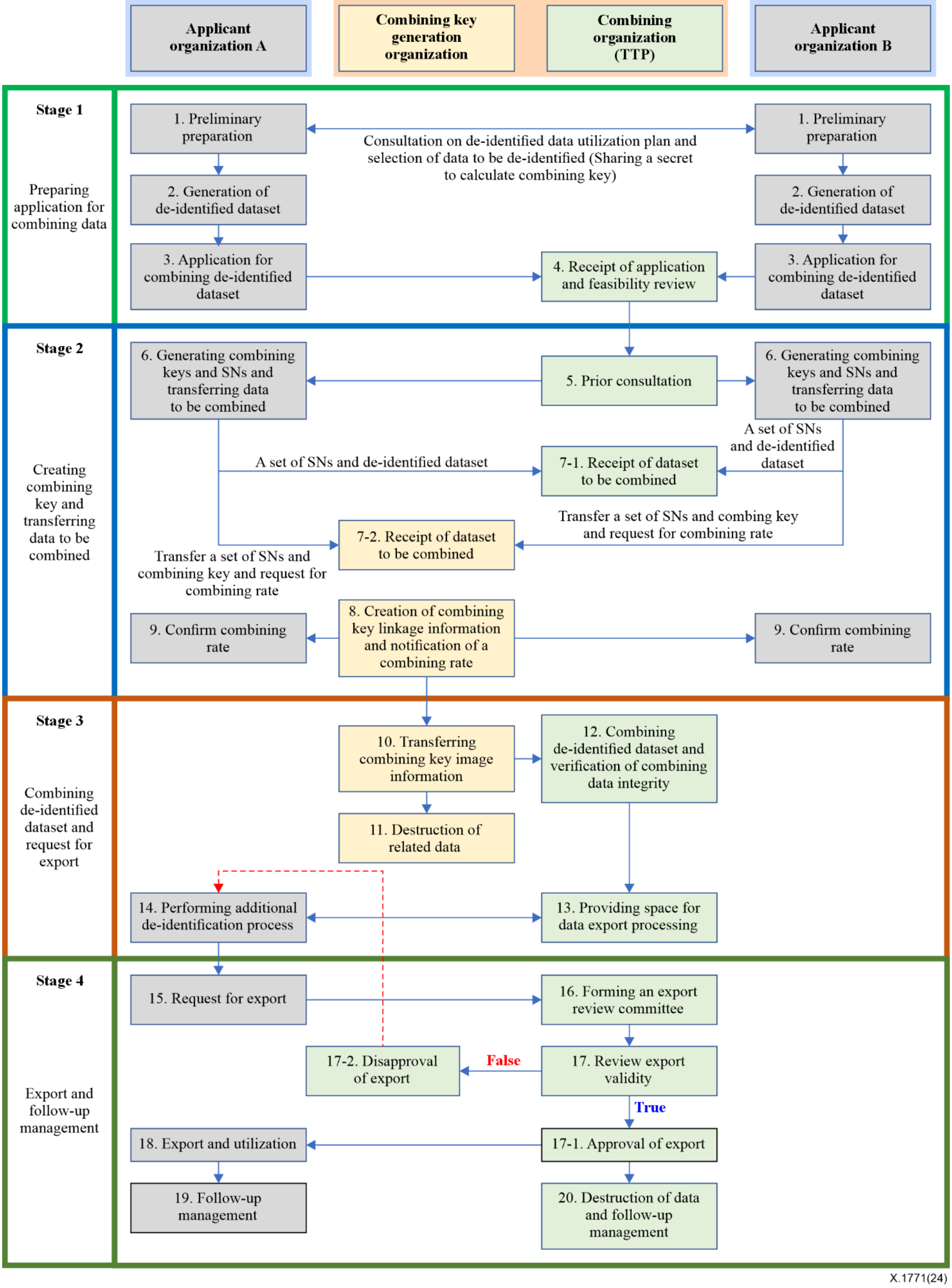
Applicant organization A
Combining key generation organization
Combining organization (TTP)
Applicant organization B
Stage 1
Preparing application for combining data
1. Preliminary preparation
Consultation on de-identified data utilization plan and selection of data to be de-identified (Sharing a secret to calculate combining key)
1. Preliminary preparation
2. Generation of de-identified dataset
2. Generation of de-identified dataset
3. Application for combining de-identified dataset
4. Receipt of application and feasibility review
3. Application for combining de-identified dataset
Stage 2
Creating combining key and transferring data to be combined
6. Generating combining keys and SNs and transferring data to be combined
5. Prior consultation
6. Generating combining keys and SNs and transferring data to be combined
A set of SNs and de-identified dataset
7-1. Receipt of dataset to be combined
A set of SNs and de-identified dataset
Transfer a set of SNs and combing key and request for combining rate
7-2. Receipt of dataset to be combined
Transfer a set of SNs and combining key and request for combining rate
9. Confirm combining rate
8. Creation of combining key linkage information and notification of a combining rate
9. Confirm combining rate
Stage 3
Combining de-identified dataset and request for export
10. Transferring combining key image information
12. Combining de-identified dataset and verification of combining data integrity
11. Destruction of related data
14. Performing additional de-identification process
13. Providing space for data export processing
Stage 4
Export and follow-up management
15. Request for export
16. Forming an export review committee
17-2. Disapproval of export
False
17. Review export validity
True
18. Export and utilization
17-1. Approval of export
19. Follow-up management
20. Destruction of data and follow-up management
X.1771(24)

Figure 8 – Procedure for combining de-identified datasets using a model with two trusted third parties

Figure 8 describes the procedure for combining de-identified datasets using the model with two trusted third parties.

In stage 1 (Preparing the application for combining datasets), there are 4 steps:

Step 1 (preliminary preparation) – Two organizations A and B conduct preliminary preparations for the combination of de-identified datasets through the mutual consultation where the selection of de-identified datasets suitable for the purpose of combination, confirmation of a series of combinations based on time, selection of items to generate a combination key, level of de-identification necessary to achieve the purpose, sharing the secret known as 'salt' to calculate the combining key, etc. are agreed. They decide which items to be used when generating a combination key among the datasets they have in common.

Step 2 (Generation of de-identified datasets) – The two organizations create a series of non-overlapping numbers (so called serial numbers) for each data record that is subject to be combined. They conduct de-identification processing for the remaining datasets except for the item for generating the serial number and the combination key.

Step 3 (Application for combining de-identified datasets) – The two organizations fill out the application for combination of de-identified datasets, prepare related documents, and submit the application for combination to the combining organization.

Step 4 (Receipt of application and feasibility review) – After receiving the applications from two organizations as a combining applicant and reviewing the feasibility of the combination, a combining organization checks the validity of application and attached documents, and if insufficient, it also requests the applicant organizations to add the missing documents.

The applicant organization should submit additional documents, etc. which are requested by the combining organization. If the combining organization recommends the level of de-identification processing, this should be reflected unless there are special rationales.

Figure 9 shows the procedure for calculating combining keys and serial numbers.

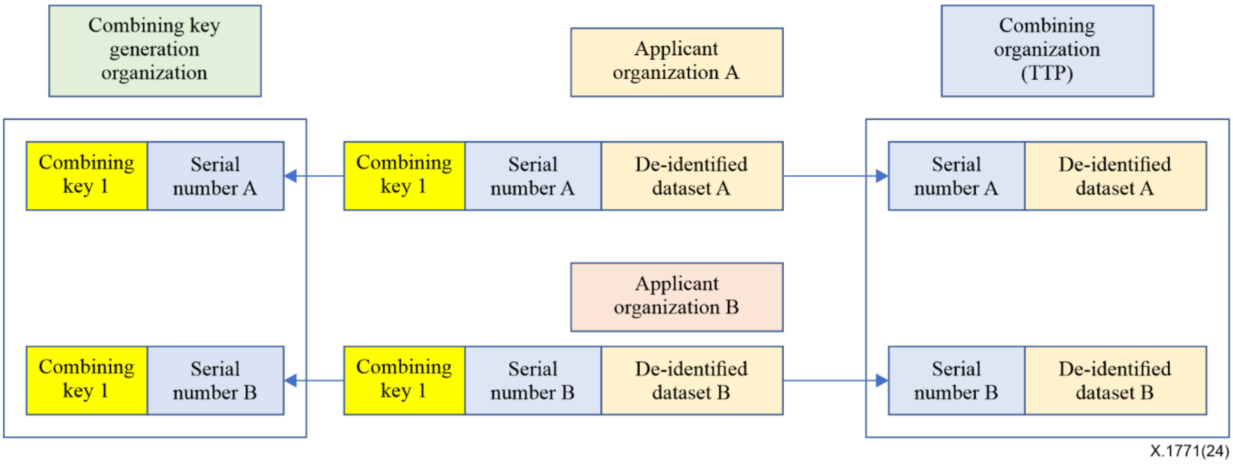

Figure 9 – Procedure for calculating combining keys and serial numbers

In stage 2 (Creating combining key and transferring dataset to be combined), there are 5 steps:

Step 5 (Prior consultation) – A combining organization and a combining key generation organization should consult in advance with the applicant organizations on matters related to a combining de-identified dataset and a combining key generation.

Step 6 (Generation of combining keys and the serial numbers (SNs) and transferring the dataset to be combined) – The applicant organization generates the combining key by applying the items agreed upon in the preliminary preparation stage and the items agreed with the combining key generation organization.

The applicant organization generates information that maps the generated combining key and the serial number (SN) and transmits them to the combining key generation organization. It also generates the combining dataset and the serial number (SN) and transmits them to the combining organization.

Step 7 (Receipt of datasets to be combined) – The combining organization and the combining key generation organization each receive the information transmitted by the applicant organizations.

Figure 10 shows the procedure for calculating combining key linkage information.

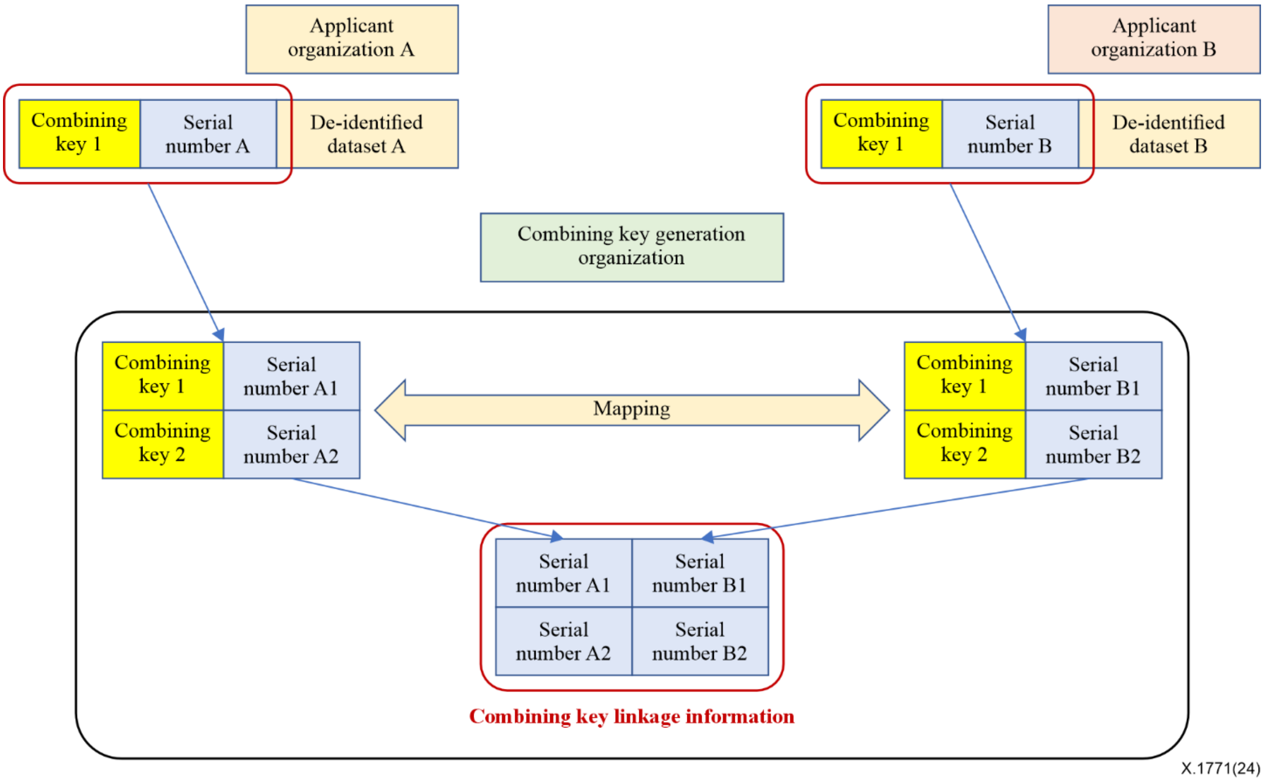

Figure 10 – Procedure for calculating combining key linkage information

Step 8 (Generation of combining key linkage information and notification of a combining rate) – The combining key organization creates the combining key linkage information consisting of 'combination key + serial number' received from the combining applicant organization. If the combining applicant organization requests a prior combining rate in the consultation stage with the combining key generation organization, the combining key organization confirms and notifies the combining rate in the process of generating the combining key linkage information.

Step 9 (Confirmation of combining rate) – The applicant organization may suspend the application for combining de-identified datasets, if necessary, after confirming the combining rate.

In stage 3 (Combining de-identified datasets and requesting for export), there are 5 steps:

Step 10 (Transferring combining key linkage information) – The combining key generation organization transmits the combining key linkage information to the combining organization.

Step 11 (Destruction of related datasets) – The combining key generation organization destroys the combining key linkage information immediately after the combining is completed (after export).

Step 12 (Combining de-identified dataset and verification of combining dataset integrity) – The combining organization performs the combining function using the de-identified datasets provided by the applicant organizations and the combining key linkage information received from the combining key generation organization. After combining, the combined dataset is verified, such as by checking the integrity of the combined de-identified dataset.

Step 13 (Providing space for dataset export processing) – The combining organization provides the export processing space so that the combining applicant organization can perform additional de-identification to export the combined de-identified dataset.

Step 14 (Performing additional de-identified processes) – The applicant (organization) shall process the combined information under an additional pseudonym or with anonymity for export in the export processing space installed within the specialized association, and shall request necessary support, such as advice, from the association specializing in association.

In stage 4 (Export and follow-up management), there are 6 steps:

Step 15 (Request for export) – The applicant organization should perform additional de-identification process, if needed, in the export processing space installed within the combining organization, and request necessary support, such as advice, from the combining organization.

Step 16 (Forming an export review committee) – When a combining applicant organization requests for export, the combining organization forms an export review committee to examine whether or not it is appropriate to export the combined de-identified dataset.

Step 17 (Review export validity and approval or disapproval of export) – If the combining applicant organization determines that it is necessary for the export review or if there is a request from the export review committee, the combining applicant organization shall submit additional datasets or attend and explain the relevant details to the export review committee.

Step 18 (Export and utilization) – The applicant organization can use the exported combined de-identified dataset for the purpose of the combination.

Step 19 (Follow-up management) – The combining applicant organizations shall monitor and conduct follow-up management.

Step 20 (Destruction of relevant dataset and follow-up management) – When the combining process is completed, the relevant history is managed and immediately deleted (combining key, etc.) except for information that is required to be separate by regulation. Follow-up management and supervision by the applicant organization is carried out, and the applicant organization shall respond to the dataset requested by the combining organization for follow-up management and supervision.

10 Controls for entities

10.1 Common technical controls applicable to all entities

The combining applicant organizations (e.g., applicant Organization A and applicant Organization B) shall take technical, managerial, and physical controls for protection of the combined de-identified dataset transferred from the trusted third party.

All communications made among relevant stakeholders shall be protected in terms of the confidentiality, integrity and availability of the combined de-identified dataset.

All entities involved in communications shall be authenticated using strong authentication methods such as multi-factor authentication.

The combining applicant organizations (e.g., applicant Organization A and applicant Organization B) shall not process the exported de-identified dataset for the purpose of identifying a specific individual and shall continuously monitor it so that it is not re-identified.

The combining key should be generated by using a hash algorithm with the input such as identifying information of a data subject and the 'salt' value (noise), in order to prevent mapping between the combining key and the identifying information of the data subject.

The combining key generation organization shall destroy all the combining key linkage information immediately after the combining process is completed (after export).

The combining organization, which is a trusted third party (TTP), shall destroy all relevant datasets for combining de-identified datasets once the combining is completed.

10.2 Controls for trusted third party

10.2.1 Technical controls for trusted party

Appropriate access controls shall be implemented to protect de-identified datasets that are to be combined.

Access to de-identified datasets shall be controlled by a strong authentication method and procedure to prove the identity of the user.

Access records to the de-identified datasets to be combined should be protected from loss, destruction, falsification, unauthorized access and unauthorized release.

10.2.2 Managerial controls for trusted third party

Only an authorized combining organization should provide combining services.

10.2.3 Physical controls

Physical controls shall be implemented for the protection of de-identified datasets by the combining organization.

Appendix I

Examples of typical de-identification techniques

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

This appendix provides a list of some typical de-identification techniques [b-ISO/IEC 20889].

- Statistical tools for de-identification techniques
- Cryptographic tools for de-identification techniques
- Suppression techniques
- Pseudonymization techniques
- Generalization techniques
- Randomization techniques
- Synthetic data

Bibliography

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

[b-ISO/IEC 27000] ISO/IEC 27000:2018, *Information technology - Security techniques - Information security management systems - Overview and vocabulary*.

[b-ISO/IEC 29100] ISO/IEC 29100:2011, *Information technology - Security techniques - Privacy framework*.

[b-NISTIR 8053] National Institute of Standards and Technology Internal Report 8053:2015, *De-Identification of Personal Information*.

[b-PIPC] Personal Information Protection Commission (2022), *Guidelines on data pseudonymization*.
https://www.pipc.go.kr/np/default/page.do?mCode=D040010000#LINK

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

Published in Switzerland
Geneva, 2024