INTERNATIONAL STANDARD

ISO/IEC 27559

First edition
2022-11

Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework

Sécurité de l'information, cybersécurité et protection de la vie privée — Cadre pour la dé-identification de données pour la protection de la vie privée

Reference number
ISO/IEC 27559:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see https://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 27, *Information security, cybersecurity and privacy protection*.

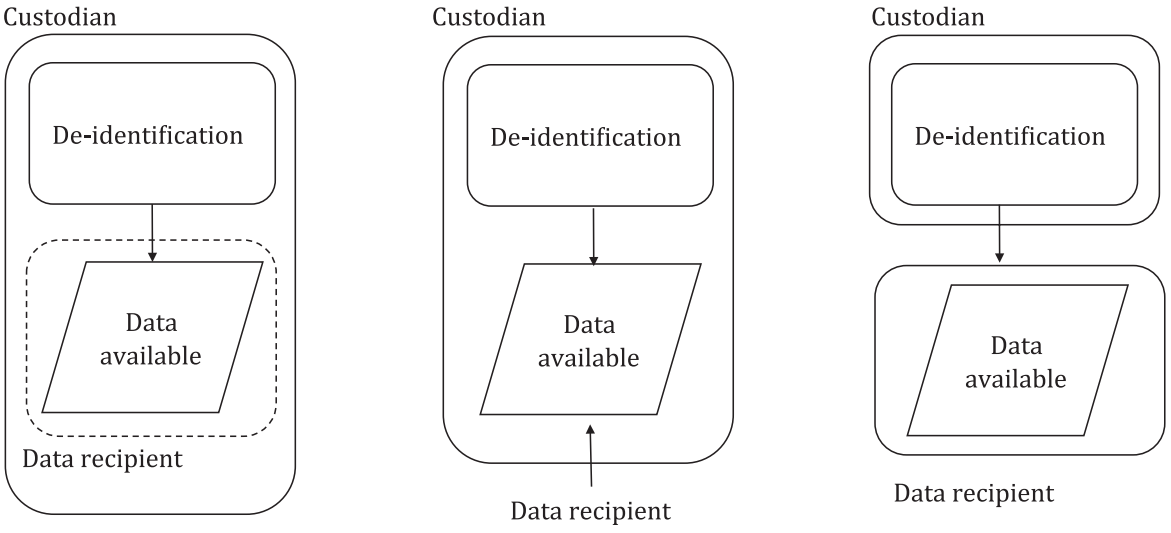
Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

De-identification is one potential means for facilitating the use of personally identifiable information (PII) in a way that does not identify or otherwise compromise the privacy of an individual or a group of individuals. The appropriate use of de-identification techniques can support compliance with regulatory requirements and relevant privacy principles. However, the term "data principal" used in this document is broader than "PII principal" and, for example, includes organizations and computers.

In almost all cases de-identification requires, at the very least, an evaluation of the additional information available to an individual or group that can inappropriately reveal or uncover PII (which is referred to as an adversary, whether a data principal is identified intentionally or not), and how they can combine it to reveal or uncover PII. In short, de-identification requires an assessment of the environment and the circumstances in which the data are made available to data recipients. This considers what additional information is available to an adversary and the possibility of attacks and motivation to re-identify. De-identification also requires an assessment of the data. This determines how the additional information available to an adversary can be used to reveal or uncover PII and the possibility of re-identification, or identity disclosure, by itself or attacks of inference.

This document provides organizations with an implementation framework to govern the appropriate use of data de-identification techniques described in ISO/IEC 20889. This de-identification framework can be applied at any point in the data lifecycle: from designing the means of data collection, the internal reuse of that data, making data available to external partners, or archival. The data recipients can therefore be internal or external to the data custodian that is implementing procedures and practices in accordance with this de-identification framework. As shown in Figure 1 a), use and reuse implies the custodian maintains oversight over the de-identified data while making it available to an internal department or functional group. Figure 1 b) shows external sharing, which implies the custodian maintains oversight over the de-identified data while making it available to an external data recipient (e.g. through a virtual access portal, or a physical data centre). Figure 1 c) shows external release, which implies the custodian transfers oversight over the de-identified data to an external data recipient. In each of these cases, the process of de-identification itself can be transferred to a third party, separate from the custodian or recipient. Written agreements with the recipient determine how data made available after de-identification can be used, in accordance with applicable laws.



a) Use and reuse **b) External sharing** **c) External release**

Figure 1 — Data availability

Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework

1 Scope

This document provides a framework for identifying and mitigating re-identification risks and risks associated with the lifecycle of de-identified data.

This document is applicable to all types and sizes of organizations, including public and private companies, government entities, and not-for-profit organizations, that are PII controllers or PII processors acting on a controller's behalf, implementing data de-identification processes for privacy enhancing purposes.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 27000, *Information technology — Security techniques — Information security management systems — Overview and vocabulary*

ISO/IEC 29100, *Information technology — Security techniques — Privacy framework*

ISO/IEC 20889, *Privacy enhancing data de-identification terminology and classification of techniques*

ISO 31000, *Risk Management — Guidelines*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 27000, ISO/IEC 29100, ISO/IEC 20889, ISO 31000 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
custodian
person or entity that has custody, control or possession of electronically stored information

[SOURCE: ISO/IEC 27050-1:2019, 3.2]

3.2
data recipient
person or organization by, with or to whom data is accessed, shared or released

3.3
adversary
individual or unit that can, whether intentionally or not, exploit potential vulnerabilities

Note 1 to entry: Adversary, attacker, intruder, snooper, and other similar terms are often used interchangeably in the de-identification literature.

3.4
threat modelling
systematic exploration technique to expose any circumstance or event having the potential to cause harm to a system in the form of destruction, *disclosure* (3.8), modification of data, or denial of service

[SOURCE: ISO/IEC/IEEE 24765:2017, 3.4290, modified — Note 1 to entry has been deleted.]

3.5
privacy impact assessment
PIA
overall process of identifying, analysing, evaluating, consulting, communicating and planning the treatment of potential privacy impacts with regard to the processing of personally identifiable information, framed within an organization's broader risk management framework

[SOURCE: ISO/IEC 29134:2017, 3.7, modified — Note 1 to entry has been deleted.]

3.6
defined population
set of elements that a dataset is drawn from that contributes to the *adversary's* (3.3) ability to identify a data principal

3.7
sample
dataset that is only a proportion of the *defined population* (3.6), such that an *adversary* (3.3) cannot be certain that any particular entity was in it

3.8
disclosure
revealing confidential or personally identifiable information from a dataset based on a vulnerability that is found or exploited

3.9
shared data
dataset in which a fixed set of entities have been granted access to the data by the custodian

3.10
released data
dataset in which the custodian no longer directly controls who has access to the data

3.11
data privacy model
approach to the application of data de-identification techniques that enables the calculation of identifiability

[SOURCE: ISO/IEC 20889:2018, 3.3, modified — The word "formal" and "measurement" have been deleted from the term and "data" added, and "re-identification risk" has been replaced by "identifiability" in the definition.]

3.12
written agreement
data sharing agreement, memorandum of understanding, data access request, contract and any other formally documented agreement

3.13
data transformation
modification of the data

3.14
de-identification governance
system of directing and controlling the de-identification process

[SOURCE: ISO/IEC 38500:2015, 2.8, modified — "the de-identification process" has been added to the definition.]

4 Symbols and abbreviated terms

PIA privacy impact assessment

PII personally identifiable information

P probability function

5 Overview

The goal of this document is to provide a principles-based framework to approach de-identification, which considers procedures, risks, and harms. A principles-based approach to de-identification is intended to be neutral on the specifics of implementation and technologies. The framework is presented in four main parts:

— Context (Clause 6): An assessment of the environment and circumstances in which the data are made available to data recipients, to determine what external information can be available to an adversary. This implies that risk can be managed through contextual controls as well (meaning the IT security controls, obligations described in written agreements, and policy and governance measures).

— Data (Clause 7): An assessment of the data, to determine how the additional information available to an adversary can be used to reveal or uncover PII. Risk can be managed by limiting what data are made available, and in what form that data will be made available (by transforming the data).

— Identifiability (Clause 8): The method of assessing identifiability is a function of context risk (the probability of an attack) and data risk (the probability of disclosure given that there is an attack). An appropriate tolerance shall be defined to ensure the identifiability is below a pre-defined tolerance level.

— Governance (Clause 9): Documented procedures and practices for the custodian to ensure the above are done consistently and effectively, now and in the future, and the preparations that are required before, during, and after de-identified data are made available.

It can be necessary to repeat the process if the resulting de-identified data does not meet acceptance criteria (by the custodian or intended data recipient), in an effort to find solutions that are acceptable to both parties. For example, the privacy and security practices for the data environment can be improved in an effort to reduce threats and improve the utility of data that are made available to the data recipients based on their operational context. The entire approach can be thought of in a somewhat linear fashion, with de-identification governance by the custodian embodying the overall processes of context assessment, data assessment, identifiability assessment and mitigation, and acceptance criteria, as shown in Figure 2. It is, however, possible to reorder elements based on implementation needs (e.g. improving efficiency and scalability for specific data flows). ISO 31000 contains guidelines on managing risk faced by organizations.

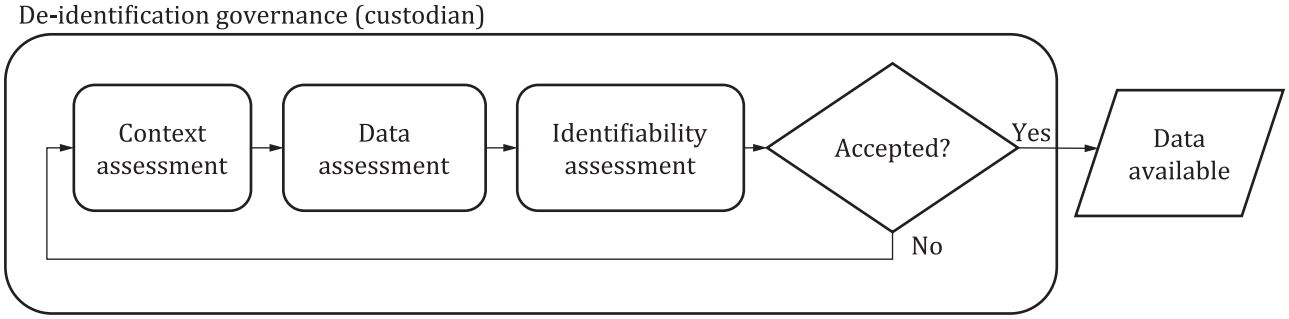


Figure 2 — De-identification framework in practice

6 Context assessment

6.1 General

The custodian shall evaluate the context in which data are being made available to a data recipient (either by providing shared access or by giving a copy of the dataset), to help properly scope the de-identification process.

NOTE Legal requirements can apply.

Determining this context involves a detailed assessment of the environment in which the data are accessed, shared or released, where the data come from and their intended use, and the circumstances under which the data are made available to a data recipient (such as levels of transparency). These elements shall be factored, in one form or another, into the method of assessing risk. For example, detailed checklists can be used to categorize risk in the data environment, and be factored into a standard risk matrix used to compare the possibility of an attack against the impact of identifying a data principal in a given context. It should be noted, however, that disclosures only occur if the attack is successful.

6.2 Threat modelling

6.2.1 General

The custodian doing the sharing or releasing, or a third party doing an assessment, shall use an objective and structured process to evaluate the environment in which the data will be accessed, shared or released. This environment includes persons and their motives (organizational and individual), other data they have access to, and their infrastructure and governance structures (including IT security controls such as those described in ISO/IEC 27002 and ISO/IEC 27701, access policies, etc.). An IT audit or assessment, even self-administered, can capture a great deal of information regarding the release environment, to help frame potential risks (in particular, potential threats).

A structured approach, often known as threat modelling, shall be used to assess the risk of an attack that would reveal or uncover PII. This includes examining what other external data sources can be available and sketching out the who, why and how of a potential disclosure. Potential threats can be:

— Deliberate: A targeted attempt to reveal or uncover PII in the data that are made available to them by an insider to the group or organization that is the data recipient.

— Accidental: A disclosure can also be unintentional, for example a data principal being recognized while a data recipient is working with the shared or released data.

— Environmental: The data can also be lost or stolen in the case where all the controls put in place have failed to prevent a data disclosure.

6.2.2 Security and privacy practices

The security and privacy practices of the data recipient will have an impact on the likelihood of a rogue employee at the data recipient's site being able to re-identify the shared data. A rogue employee can choose not to abide by a contract in the absence of strong mitigating controls. The security and privacy practices can also determine the likelihood of an outsider gaining access to the shared data, either directly or by compromising an insider's credentials.

An evaluation of mitigating controls shall be detailed and evidence based.

NOTE Professional, international, and government regulations, standards, and policies can apply, including ISO/IEC 27002 and ISO/IEC 27701, where appropriate.

Using a standardized approach also ensures consistency, not only for a single organization that is sharing data, but across organizations.

In order to avoid inappropriate or excessive burdens on the data provider or recipient, the evaluation can take into account third party audits and relevant certifications, as well as re-using prior analyses.

6.2.3 Motives and capacity to re-identify

A recipient's motives to reveal or uncover PII can be controlled in part by training, awareness, and obligations described in written agreements, including processes and terms described in ISO/IEC 23751, provided they are enforceable (through legal mechanisms and by refusing to share or release additional data). Obligations in written agreements can include:

— delivery of training on disclosure risks to individuals with access to de-identified data;

— regular reminders of their obligations to uphold data privacy and security policies;

— prohibiting attempts to identify data principals in the data made available to data recipients, or linking data that would extend the profiles of data principals (thereby increasing the risk of disclosing PII) without express permission;

— allowing for spot checks or full audits (possibly by a third party) that ensure compliance with the stated terms of the agreement;

— prohibiting the sharing with any other data recipient without express permission;

— defining the environment in which the data are accessed, shared and released (the infrastructure and governance structures that are expected to be in place);

— defining acceptable use cases and any expectations of how the data are used, or how use cases are evaluated to ensure its use is for purposes that are deemed appropriate.

NOTE It is possible that the enforcement of agreement obligations is not as effective when publishing as open data.

In planning the above obligations, the custodian should know where the data have come from, where they are going, and what the intended uses are for the data being accessed, shared or released. This will help situate the data release and the environment in which they will be made available to data recipients, so that appropriate measures can be taken to reduce the risk of potential disclosures of PII.

6.3 Transparency and impact assessment

6.3.1 General

The custodian shall assess the impact of disclosure, which decides tolerance when identifiability is evaluated, and can include:

— Being as transparent as possible and engaging with stakeholders where practicable. This way expectations of privacy are better understood by the entities involved.

— Whether the data are highly sensitive and intimate, come from vulnerable populations, or can reveal sensitive attributes or be stigmatizing.

— Potential injury or harm to data principals that can arise due to inappropriate processing.

— The trustworthiness of the data recipients (e.g. data sharing agreement in place, history of partnership, incentives to remain an ongoing partner).

— How the purposes for sharing and intended uses are in line with the interest of data principals.

— The extent to which data recipients will thoroughly consider their potential uses of data (for example, the use of privacy impact assessments) and act accordingly.

— Any potential breach of legal principles or fundamental rights.

6.3.2 Transparency of actions and stakeholder engagement

The custodian should explain simply and clearly its data collection practices, how it reuses data with a description of its rationale, and be open to feedback or seek the views of stakeholders on its data sharing activities with the goal of understanding, and where appropriate, addressing their concerns. Trust requires openness. Meaningful discussions with stakeholders can help to establish a reasonable balance of risk and benefits.

The custodian can consider comparing how similar organizations in its sector are releasing or sharing data, and whether any concerns have been raised about their practices. Surveys and focus group work on the data principals' views of data release, sharing and reuse (e.g., by industry associations) can also be considered to help inform decision making.

6.3.3 Privacy-related harms

The custodian should consider the potential privacy-related harms that can result from data being made available, including potential injury or harm due to inappropriate processing. For example, the custodian may:

a) evaluate whether the data are highly sensitive and personal; and also

b) consider use cases that may:

   1) reveal sensitive attributes,

   2) be stigmatizing, or

   3) support decisions that adversely affect data principals.

A custodian should conduct privacy impact assessments (PIA) in order to identify privacy-related risks and the proper controls to mitigate them. The PIA process may be shared and explained to increase transparency. It is a process that begins at the earliest possible stages of an initiative, when there are