INTERNATIONAL STANDARD

ISO 25237

First edition
2017-01

Health informatics — Pseudonymization

Informatique de santé — Pseudonymisation

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

Reference number
ISO 25237:2017(E)

© ISO 2017

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

COPYRIGHT PROTECTED DOCUMENT

© ISO 2017, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

© ISO 2017 – All rights reserved

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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 ISO documents 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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO 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).

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

For an explanation on 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 the following URL: www.iso.org/iso/foreword.html.

The committee responsible for this document is ISO/TC 215, *Health informatics*.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

Introduction

Pseudonymization is recognized as an important method for privacy protection of personal health information. Such services may be used nationally, as well as for trans-border communication.

Application areas include, but are not limited to:

- indirect use of clinical data (e.g. research);
- clinical trials and post-marketing surveillance;
- pseudonymous care;
- patient identification systems;
- public health monitoring and assessment;
- confidential patient-safety reporting (e.g. adverse drug effects);
- comparative quality indicator reporting;
- peer review;
- consumer groups;
- field service.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

This document provides a conceptual model of the problem areas, requirements for trustworthy practices, and specifications to support the planning and implementation of pseudonymization services.

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

The specification of a general workflow, together with a policy for trustworthy operations, serve both as a general guide for implementers but also for quality assurance purposes, assisting users of the pseudonymization services to determine their trust in the services provided. This guide will serve to educate organizations so they can perform pseudonymization services themselves with sufficient proficiency to achieve the desired degree of quality and risk reduction.

© ISO 2017 – All rights reserved

INTERNATIONAL STANDARD **ISO 25237:2017(E)**

Health informatics — Pseudonymization

1 Scope

This document contains principles and requirements for privacy protection using pseudonymization services for the protection of personal health information. This document is applicable to organizations who wish to undertake pseudonymization processes for themselves or to organizations who make a claim of trustworthiness for operations engaged in pseudonymization services.

This document

— defines one basic concept for pseudonymization (see Clause 5),

— defines one basic methodology for pseudonymization services including organizational, as well as technical aspects (see Clause 6),

— specifies a policy framework and minimal requirements for controlled re-identification (see Clause 7),

— gives an overview of different use cases for pseudonymization that can be both reversible and irreversible (see Annex A),

— gives a guide to risk assessment for re-identification (see Annex B),

— provides an example of a system that uses de-identification (see Annex C),

— provides informative requirements to an interoperability to pseudonymization services (see Annex D), and

— specifies a policy framework and minimal requirements for trustworthy practices for the operations of a pseudonymization service (see Annex E).

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

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 27799, *Health informatics — Information security management in health using ISO/IEC 27002*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

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

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

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

3.1
access control
means of ensuring that the resources of a data processing system can be accessed only by authorized entities in authorized ways

[SOURCE: ISO/IEC 2382:2015, 2126294]

3.2

anonymization

process by which *personal data* (3.37) is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party

Note 1 to entry: The concept is absolute, and in practice, it may be difficult to obtain.

[SOURCE: ISO/IEC 29100:2011, 2.2, modified.]

3.3

anonymized data

data (3.14) that has been produced as the output of an *anonymization* (3.2) process

[SOURCE: ISO/IEC 29100:2011, 2.3, modified.]

3.4

anonymous identifier

identifier (3.27) of a person which does not allow the *identification* (3.26) of the *natural person* (3.34)

3.5

authentication

assurance of the claimed identity

3.6

attacker

person deliberately exploiting vulnerabilities in technical and non-technical security controls in order to steal or compromise information systems and networks, or to compromise availability to legitimate users of information system and network resources

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[SOURCE: ISO/IEC 27033-1:2015, 3.3]

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

3.7

ciphertext

data (3.14) produced through the use of encryption, the semantic content of which is not available without the use of cryptographic techniques

[SOURCE: ISO/IEC 2382:2015, 2126285]

3.8

confidentiality

property that *information* (3.29) is not made available or disclosed to unauthorized individuals, entities or processes

[SOURCE: ISO 7498-2:1989, 3.3.16]

3.9

content-encryption key

cryptographic key used to encrypt the content of a communication

3.10

controller

natural or legal person, public authority, agency or any other body which, alone or jointly with others, determines the purposes and means of the *processing of personal data* (3.40)

3.11

cryptography

discipline which embodies principles, means and methods for the transformation of *data* (3.14) in order to hide its information content, prevent its undetected modification and/or prevent its unauthorized use

[SOURCE: ISO 7498-2:1989, 3.3.20]

3.12
cryptographic algorithm
<cipher> method for the transformation of *data* (3.14) in order to hide its information content, prevent its undetected modification and/or prevent its unauthorized use

3.13
cryptographic key management
key management
generation, storage, distribution, deletion, archiving and application of *keys* (3.31) in accordance with a *security policy* (3.46)

[SOURCE: ISO 7498-2:1989, 3.3.33]

3.14
data
reinterpretable representation of *information* (3.29) in a formalized manner suitable for communication, interpretation or processing

Note 1 to entry: Data can be processed by humans or by automatic means.

[SOURCE: ISO/IEC 2382:2015, 2121272]

3.15
data integrity
property that *data* (3.14) has not been altered or destroyed in an unauthorized manner

[SOURCE: ISO 7498-2:1989, 3.3.21]

3.16
data linking
matching and combining *data* (3.14) from multiple databases

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

3.17
data protection
technical and social regimen for negotiating, managing and ensuring informational *privacy* (3.39), and security

3.18
data subject
person to whom *data* (3.14) refer

3.19
decryption
process of converting encrypted *data* (3.14) back into its original form so it can be understood

3.20
de-identification
general term for any process of reducing the association between a set of identifying *data* (3.14) and the *data subject* (3.18)

3.21
directly identifying data
data (3.14) that directly identifies a single individual

Note 1 to entry: Direct identifiers are those data that can be used to identify a person without additional information or with cross-linking through other information that is in the public domain.

3.22
disclosure
divulging of, or provision of access to, *data* (3.14)

Note 1 to entry: Whether the recipient actually looks at the data, takes them into knowledge or retains them, is irrelevant to whether disclosure has occurred.

3.23
encryption
process of converting *information* (3.29) or *data* (3.14) into a cipher or code

3.24
healthcare identifier
subject of care identifier
identifier (3.27) of a person for primary use by a healthcare system

3.25
identifiable person
one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity

[SOURCE: Directive 95/46/EC]

3.26
identification
process of using claimed or observed attributes of an entity to single out the entity among other entities in a set of identities

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Note 1 to entry: The identification of an entity within a certain context enables another entity to distinguish between the entities with which it interacts.

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

3.27
identifier
information (3.29) used to claim an identity, before a potential corroboration by a corresponding authenticator

[SOURCE: ENV 13608-1:2000, 3.44]

3.28
indirectly identifying data
data (3.14) that can identify a single person only when used together with other indirectly identifying data

Note 1 to entry: Indirect identifiers can reduce the population to which the person belongs, possibly down to one if used in combination.

EXAMPLE Postcode, sex, age, date of birth.

3.29
information
knowledge concerning objects that within a certain context has a particular meaning

[SOURCE: ISO/IEC 2382:2015, 2121271, modified.]

3.30
irreversibility
situation when, for any passage from identifiable to pseudonymous, it is computationally unfeasible to trace back to the original *identifier* (3.27) from the *pseudonym* (3.43)

© ISO 2017 – All rights reserved

3.31
key
sequence of symbols which controls the operations of *encryption* (3.23) and *decryption* (3.19)

[SOURCE: ISO 7498-2:1989, 3.3.32]

3.32
linkage of information objects
process allowing a logical association to be established between different information objects

3.33
longitudinal or lifetime personal health record
permanent, coordinated record of significant information, in chronological sequence

Note 1 to entry: It may include all historical data collected or be retrieved as a user designated synopsis of significant demographic, genetic, clinical and environmental facts and events maintained within an automated system.

[SOURCE: ISO/TR 21089:2004, 3.61, modified]

3.34
natural person
real human being as opposed to a legal person which may be a private or public organization

3.35
person identification
process for establishing an association between an information object and a physical person

iTeh STANDARD PREVIEW
(standards.iteh.ai)

3.36
personal identifier
information with the purpose of uniquely identifying a person within a given context

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

3.37
personal data
information relating to an identified or identifiable *natural person* (3.34) (“data subject”)

[SOURCE: Directive 95/46/EC]

3.38
primary use of personal data
uses and *disclosures* (3.22) that are intended for the *data* (3.14) collected

3.39
privacy
freedom from intrusion into the private life or affairs of an individual when that intrusion results from undue or illegal gathering and use of *data* (3.14) about that individual

[SOURCE: ISO/IEC 2382:2015, 2126263]

3.40
processing of personal data
operation or set of operations that is performed upon *personal data* (3.37), whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction

[SOURCE: Directive 95/46/EC]

3.41
processor
natural or legal person, public authority, agency or any other body that processes *personal data* (3.37) on behalf of the *controller* (3.10)

Note 1 to entry: See Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

3.42
pseudonymization
particular type of *de-identification* (3.20) that both removes the association with a *data subject* (3.18) and adds an association between a particular set of characteristics relating to the data subject and one or more *pseudonyms* (3.43)

3.43
pseudonym
personal identifier (3.36) that is different from the normally used personal identifier and is used with pseudonymized data to provide dataset coherence linking all the information about a subject, without disclosing the real world person identity.

Note 1 to entry: This may be either derived from the normally used personal identifier in a reversible or irreversible way or be totally unrelated.

Note 2 to entry: Pseudonym is usually restricted to mean an identifier that does not allow the direct derivation of the normal personal identifier. Such pseudonymous information is thus functionally anonymous. A trusted third party may be able to obtain the normal personal identifier from the pseudonym.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

3.44
recipient
natural or legal person, public authority, agency or any other body to whom *data* (3.14) are disclosed

3.45
secondary use of personal data
uses and *disclosures* (3.22) that are different than the initial intended use for the *data* (3.14) collected

3.46
security policy
plan or course of action adopted for providing computer security

[SOURCE: ISO/IEC 2382:2015, 2126246]

3.47
trusted third party
security authority, or its agent, trusted by other entities with respect to security-related activities

[SOURCE: ISO/IEC 18014-1:2008, 3.20]

4 Abbreviated terms

DICOM	Digital Imaging and Communication in Medicine
HIPAA	Health Insurance Portability and Accountability Act
HIS	Health Information System
HIV	Human Immunodeficiency Virus
IP	Internet Protocol
VoV	Victim of Violence use

5 Requirements for privacy protection of identities in healthcare

5.1 Objectives of privacy protection

The objective of privacy protection as part of the confidentiality objective of security is to prevent the unauthorized or unwanted disclosure of information about a person which may further influence legal, organizational and financial risk factors. Privacy protection is a subdomain of generic privacy protection that, by definition, includes other privacy sensitive entities such as organizations. As privacy is the best regulated and pervasive one, this conceptual model focuses on privacy. Protective solutions designed for privacy can also be transposed for the privacy protection of other entities. This may be useful in countries where the privacy of entities or organizations is regulated by law.

There are two objectives in the protection of personal data; one that is the protection of personal data in interaction with on-line applications (e.g. web browsing) and at the other is the protection of collected personal data in databases. This document will restrict itself to the latter objective.

Data can be extracted from databases. The objective is to reduce the risk that the identities of the data subjects are disclosed. Researchers work with “cases”, longitudinal histories of patients collected in time and/or from different sources. For the aggregation of various data elements into the cases, it is, however, necessary to use a technique that enables aggregations without endangering the privacy of the data subjects whose data are being aggregated. This can be achieved by pseudonymization of the data.

De-identification is used to reduce privacy risks in a wide variety of situations.

Extreme de-identification is used for educational materials that will be made widely public, yet should convey enough detail to be useful for medical education purposes (there is an IHE profile for automation assistance for performing this kind of de-identification. Much of the process is customized to the individual patient and educational purpose).

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Public health uses de-identified databases to track and understand diseases.

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

Clinical trials use de-identification both to protect privacy and to avoid subconscious bias by removing other information such as whether the patient received a placebo or an experimental drug.

Slight de-identification is used in many clinical reviews, where the reviewers are kept ignorant of the treating physician, hospital, patient, etc. both to reduce privacy risks and to remove subconscious biases. This kind of de-identification only prevents incidental disclosure to reviewers. An intentional effort will easily discover the patient identity, etc.

When undertaking production of workload statistics or workload analysis within hospitals or of treatments provided against contracts with commissioners or purchasers of health care services, it is necessary to be able to separate individual patients without the need to know who the individual patients are. This is an example of the use of de-identification within a business setting.

The process of risk stratification (of re-hospitalization, for example) can be undertaken by using records from primary and secondary care services for patients. The records are de-identified for the analysis, but where the patients that are indicated as being of high risk, these patients can be re-identified by an appropriate clinician to enable follow-up interventions. For details on the healthcare pseudonymizaton, see Annex A.

5.2 General

De-identification is the general term for any process of reducing the association between a set of identifying data and the data subject with one or more intended use of the resulting data-set. Pseudonymization is a subcategory of de-identification. The pseudonym is the means by which pseudonymized data are linked to the same person or information systems without revealing the identity of the person. De-identification inherently can limit the utility of the resulting data. Pseudonymization can be performed with or without the possibility of re-identifying the subject of the data (reversible or irreversible pseudonymization). There are several use case scenarios in healthcare for pseudonymization with particular applicability in increasing electronic processing of patient data,

© ISO 2017 – All rights reserved

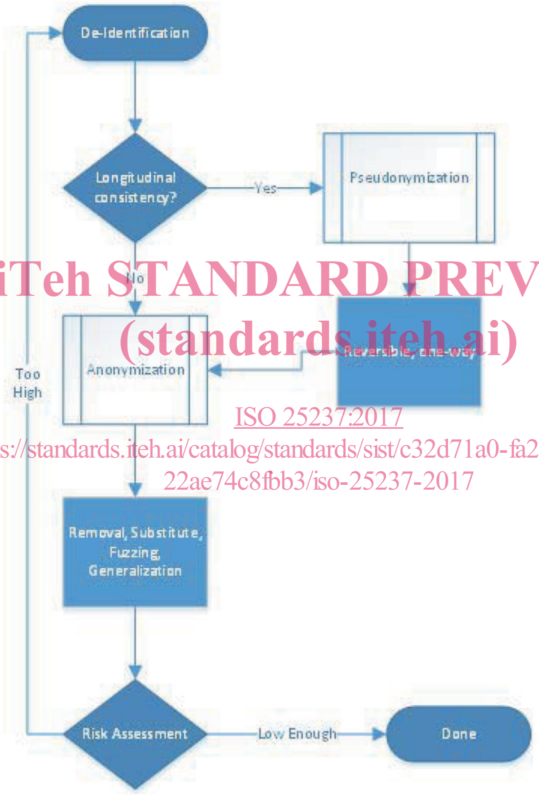
together with increasing patient expectations for privacy protection. Several examples of these are provided in Annex A.

It is important to note that as long as there are any pseudonymized data, there is some risk of unauthorized re-identification. This is not unlike encryption, in that brute force can crack encryption, but the objective is to make it so difficult that the cost is prohibitive. There is less experience with de-identification than encryption so the risks are not as well understood.

5.3 De-identification as a process to reduce risk

5.3.1 General

The de-identification process should consider the security and privacy controls that will manage the resulting data-set. It is rare to lower the risk so much that the data-set needs no ongoing security controls.



iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

Figure 1 — Visualization of the de-identification process

Figure 1 is an informative diagram of a visualization of this de-identification process. This shows that the topmost concept is de-identification, as a process. This process utilizes sub-processes: pseudonymization and/or anonymization. These sub-processes use various tools that are specific to the type of data element they operate on, and the method of risk reduction.

The starting state is that zero data are allowed to pass through the system. Each element should be justified by the intended use of the resulting data-set. This intended use of the data-set greatly affects the de-identification process.

5.3.2 Pseudonymization

De-identification might leverage pseudonymization where longitudinal consistency is needed. This might be to keep a bunch of records together that should be associated with each other, where without this longitudinal consistency, they might get disassociated. This is useful to keep all of the records

© ISO 2017 – All rights reserved

for a patient together, under a pseudonym. This also can be used to assure that each time data are extracted into a de-identified set that new entries are also associated with the same pseudonym. In pseudonymization, the algorithm used might be intentionally reversible or intentionally not-reversible. A reversible scheme might be a secret lookup-table that where authorized can be used to discover the original identity. In a non-reversible scheme, a temporary table might be used during the process, but is destroyed when the process completes.

5.3.3 Anonymization

Anonymization is the process and set of tools used where no longitudinal consistency is needed. The anonymization process is also used where pseudonymization has been used to address the remaining data attributes. Anonymization utilizes tools like redaction, removal, blanking, substitution, randomization, shifting, skewing, truncation, grouping, etc. Anonymization can lead to a reduced possibility of linkage.

Each element allowed to pass should be justified. Each element should present the minimal risk, given the intended use of the resulting data-set. Thus, where the intended use of the resulting data-set does not require fine-grain codes, a grouping of codes might be used.

5.3.4 Direct and indirect identifiers

De-identification process addresses three kinds of data: direct identifiers, which by themselves identify the patient; indirect identifiers, which provide correlation when used with other indirect or external knowledge; and non-identifying data, the rest of the data.

Usually, a de-identification process is applied to a data-set, made up of entries that have many attributes. For example, a spreadsheet made up of rows of data organized by column.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

The de-identification process, including pseudonymization and anonymization, are applied to all the data. Pseudonymization generally are used against direct identifiers, but might be used against indirect identifiers, as appropriate to reduce risk while maintaining the longitudinal needs of the intended use of the resulting data-set. Anonymization tools are used against all forms of data, as appropriate to reduce risk.

ISO 25237:2017
https://standards.iteh.ai/catalog/standards/sist/c32d71a0-fa2f-44b6-8ab4-22ae74c8fbb3/iso-25237-2017

5.4 Privacy protection of entities

5.4.1 Personal data versus de-identified data

5.4.1.1 Definition of personal data

According to Reference [18], "personal data" shall mean any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

This concept is addressed in other national legislation with consideration for the same principles found in this definition (e.g. HIPAA).