DRAFT INTERNATIONAL STANDARD

ISO/DIS 25237

ISO/TC **215**	Secretariat: **ANSI**
Voting begins on: **2015-09-03**	Voting terminates on: **2015-12-03**

Health informatics — Pseudonymisation

Informatique de santé — Pseudonymization

ICS: 35.240.80

ISO/CEN PARALLEL PROCESSING

This draft has been developed within the International Organization for Standardization (ISO), and processed under the **ISO lead** mode of collaboration as defined in the Vienna Agreement.

This draft is hereby submitted to the ISO member bodies and to the CEN member bodies for a parallel five month enquiry.

Should this draft be accepted, a final draft, established on the basis of comments received, will be submitted to a parallel two-month approval vote in ISO and formal vote in CEN.

To expedite distribution, this document is circulated as received from the committee secretariat. ISO Central Secretariat work of editing and text composition will be undertaken at publication stage.

Reference number
ISO/DIS 25237:2015(E)

COPYRIGHT PROTECTED DOCUMENT

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

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 WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information.

The committee responsible for this document is ISO/TC 215, *Health informatics*.ISO/TS 25237 was prepared by Technical Committee ISO/TC 215, *Healthcare informatics*.

Introduction

Pseudonymization is recognised 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.

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

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 education organizations so they can perform pseudonymization services themselves with sufficient proficiency to achieve the desired degree of quality and risk reduction.

Health informatics — Pseudonymization

1 Scope

This Technical Specification contains principles and requirements for privacy protection using pseudonymization services for the protection of personal health information. This technical specification 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 Technical Specification:

— defines one basic concept for pseudonymization;

— gives an overview of different use cases for pseudonymization that can be both reversible and irreversible;

— defines one basic methodology for pseudonymization services including organizational as well as technical aspects;

— gives a guide to risk assessment for re-identification;

— specifies a policy framework and minimal requirements for trustworthy practices for the operations of a pseudonymization service;

— specifies a policy framework and minimal requirements for controlled re-identification;

2 Normative references

The following referenced documents are indispensable for the application 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*

IHE Healthcare De-Identification Handbook: 2014

3 Terms and definitions

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

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

[ISO/IEC 2382-8:1998, definition 08.04.01]

3.2
anonymization
process by which personal data 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

ICS 35.240.80
Price based on 57 pages

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

NOTE: Adapted from ISO/IEC 29100:2011 Information technology — Security techniques — Privacy framework

3.3
anonymized data

data that has been produced as the output of an anonymization process

NOTE: Adapted from ISO/IEC 29100:2011 Information technology — Security techniques — Privacy framework

3.4
anonymous identifier
identifier of a person which does not allow the identification of the natural person -

3.5
authentication
assurance of the claimed identity

3.6
attacker
person seeking to exploit potential vulnerabilities of a biometric system
[ISO/IEC 19792:2009(en)

3.7
ciphertext
data produced through the use of encryption, the semantic content of which is not available without the use of cryptographic techniques

[ISO/IEC 2382-8:1998, definition 08-03-8]

3.8
confidentiality
property that information is not made available or disclosed to unauthorized individuals, entities or processes

[ISO 7498-2:1989, definition 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

3.11
cryptography
discipline which embodies principles, means and methods for the transformation of data in order to hide its information content, prevent its undetected modification and/or prevent its unauthorized use

[ISO 7498-2:1989, definition 3.3.20]

3.12
cryptographic algorithm
⟨cipher⟩ method for the transformation of data in order to hide its information content, prevent its undetected modification and/or prevent its unauthorized use

3.13
key management
cryptographic key management
generation, storage, distribution, deletion, archiving and application of keys in accordance with a **security policy** (3.43)

[ISO 7498-2:1989, definition 3.3.33]

3.14
data integrity
property that data have not been altered or destroyed in an unauthorized manner

[ISO 7498-2:1989, definition 3.3.21]

3.15
data linking
matching and combining data from multiple databases

3.16
data protection
technical and social regimen for negotiating, managing and ensuring informational privacy, and security

3.17
data-subjects
persons to whom data refer

3.18
decipherment
decryption
process of obtaining, from a ciphertext, the original corresponding data

[ISO/IEC 2382-8:1998, definition 08-03-04]

NOTE A ciphertext can be enciphered a second time, in which case a single decipherment does not produce the original plaintext.

3.19
de-identification
general term for any process of reducing the association between a set of identifying data and the data subject

3.20
Directly Identifying data
data that directly identifies a single individual

NOTE 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.

3.21
disclosure
divulging of, or provision of access to, data

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

ICS 35.240.80

Price based on 57 pages

3.22
encipherment encryption
cryptographic transformation of data to produce **ciphertext** (3.6)

[ISO 7498-2:1989, definition 3.3.27]

NOTE See **cryptography** (3.10).

3.23
subject of care identifier
healthcare identifier
identifier of a person for primary use by a healthcare system

3.24
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

[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]

3.25
identification
process of using claimed or observed attributes of an entity to single out the entity among other entities in a set of identities

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

3.26
identifier
information used to claim an identity, before a potential corroboration by a corresponding authenticator (as used in this document)

[ENV 13608-1]

3.27
indirectly identifying data
data that can identify a single person only when used together with other indirectly identifying data

NOTE 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.

3.28
information
knowledge concerning objects that within a certain context has a particular meaning
[ISO/IEC 2382-1:1993]

NEW TERM
data
reinterpretable representation of information in a formalized manner suitable for communication, interpretation or processing
NOTE: Data can be processed by humans or by automatic means.
[ISO/IEC 2382-1:1993]

3.29
irreversibility
situation when, for any passage from identifiable to pseudonymous, it is computationally unfeasible to trace back to the original identifier from the pseudonym

3.30
key
sequence of symbols which controls the operations of **encipherment** (3.21) and **decipherment** (3.17)

[ISO 7498-2:1989, definition 3.3.32]

3.31
linkage of information objects
process allowing a logical association to be established between different information objects

3.32
natural person
Real human being as opposed to a legal person which may be a private or public organisation

3.33
person identification
process for establishing an association between an information object and a physical person

3.34
personal identifier
information with the purpose of uniquely identifying a person within a given context

3.35
personal data
any information relating to an identified or identifiable natural person (“data subject”)

[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]

3.36
primary use of personal data
use of personal data for delivering healthcare

3.37
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 about that individual

[ISO/IEC 2382-8:1998, definition 08-01-23]

3.38
processing of personal data
any operation or set of operations that is performed upon personal data, 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

[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]

ICS 35.240.80
Price based on 57 pages

3.39
processor
natural or legal person, public authority, agency or any other body that processes personal data on behalf of the controller

[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]

3.40
pseudonymization
particular type of de-identification that both removes the association with a data subject and adds an association between a particular set of characteristics relating to the data subject and one or more pseudonyms

3.41
pseudonym
personal identifier 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 This may be either derived from the normally used personal identifier in a reversible or irreversible way, or be totally unrelated.

NOTE 2 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.

3.42
recipient
natural or legal person, public authority, agency or any other body to whom data are disclosed

3.43
~~secondary~~ indirect use of personal data
Indirect uses are those that are different than the initial intended use for the data collected.

3.44
security policy
plan or course of action adopted for providing computer security

[ISO/IEC 2382-8:1998, definition 08-01-06]

3.45
trusted third party
security authority, or its agent, trusted by other entities with respect to security-related activities

[ISO/IEC 18014-1:2008]

4 Symbols (and abbreviated terms)

HIPAA Health Insurance Portability and Accountability Act

HIS Hospital Information System

HIV Human Immunodeficiency Virus

IP Internet Protocol

VoV Victim of Violence

5 Requirements for privacy protection of identities in healthcare

5.1.1 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 Technical Specification 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 must 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.)

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

- 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-hospitalisation 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.

5.1.2 General

De-identification is the general term for any process of removing 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. Pseudonymization can be

ICS 35.240.80
Price based on 57 pages