INTERNATIONAL STANDARD

ISO/IEC 20889

First edition
2018-11

Privacy enhancing data de-identification terminology and classification of techniques

Terminologie et classification des techniques de dé-identification de données pour la protection de la vie privée

Reference number
ISO/IEC 20889:2018(E)

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

Published in Switzerland

Contents

Page

Foreword ..... **v**
Introduction ..... **vi**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Symbols and abbreviated terms **5**
5 Overview **6**
6 Technical model and terminology **6**
7 Re-identification **8**
7.1 General 8
7.2 Re-identification attacks 8
8 Usefulness of de-identified data **10**
9 De-identification techniques **10**
9.1 Statistical tools 10
9.1.1 General 10
9.1.2 Sampling 10
9.1.3 Aggregation 11
9.2 Cryptographic tools 11
9.2.1 General 11
9.2.2 Deterministic encryption 11
9.2.3 Order-preserving encryption 12
9.2.4 Format-preserving encryption 12
9.2.5 Homomorphic encryption 13
9.2.6 Homomorphic secret sharing 13
9.3 Suppression techniques 14
9.3.1 General 14
9.3.2 Masking 14
9.3.3 Local suppression 15
9.3.4 Record suppression 15
9.4 Pseudonymization techniques 15
9.4.1 General 15
9.4.2 Selection of attributes 15
9.4.3 Creation of pseudonyms 16
9.5 Anatomization 17
9.6 Generalization techniques 17
9.6.1 General 17
9.6.2 Rounding 18
9.6.3 Top and bottom coding 18
9.6.4 Combining a set of attributes into a single attribute 18
9.6.5 Local generalization 18
9.7 Randomization techniques 18
9.7.1 General 18
9.7.2 Noise addition 19
9.7.3 Permutation 19
9.7.4 Microaggregation 19
9.8 Synthetic data 20
10 Formal privacy measurement models **20**
10.1 General 20
10.2 ***K*-anonymity model** 20
10.2.1 General 20

10.2.2 ***L*-diversity** 21
10.2.3 ***T*-closeness** 21
10.3 Differential privacy model 21
10.3.1 General 21
10.3.2 Server model 22
10.3.3 Local model 22
10.3.4 Key considerations for a Differentially Private System 23
10.4 Linear sensitivity model 24
10.4.1 General 24
10.4.2 Threshold rule 24
10.4.3 Dominance rule 25
10.4.4 Ambiguity rule 25

11 General principles for application of de-identification techniques 25
11.1 General 25
11.2 Sampling considerations 25
11.3 Aggregated vs. microdata 26
11.4 Classification of attributes 26
11.5 Handling of direct identifiers 26
11.6 Handling of remaining attributes 26
11.7 Privacy guarantee models 27

12 Additional technical or organizational measures 27
12.1 General 27
12.2 Data flow scenarios 27
12.3 Access to de-identified data 28
12.4 Controlled re-identification 28

Annex A (informative) **Summary of de-identification tools and techniques 29**

Annex B (informative) **Prior art terminology 31**

Annex C (informative) **De-identification of free-form text 34**

Annex D (informative) **Normalization of structured data 37**

Annex E (informative) **Overview of approaches to formal privacy measurement models 38**

Bibliography 43

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).

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 http://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.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 27, *IT Security techniques*.

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.

Introduction

It is well-established that major benefits can be derived from processing electronically stored data, including so-called "big data". However, where this data includes personally identifiable information (PII), as is often the case, processing this data needs to comply with applicable personal data protection principles. The appropriate use of de-identification techniques is an important component of measures to enable the exploitation of the benefits of data processing while maintaining compliance with the relevant ISO/IEC 29100 privacy principles.

The immediate relevance of this document is to personal data protection of natural persons (i.e. PII principals), but the term "data principal", defined and used in this document, is broader than "PII principal" and, for example, includes organizations and computers.

This document focuses on commonly used techniques for de-identification of structured datasets as well as on datasets containing information about data principals that can be represented logically in the form of a table. In particular, the techniques are applicable to datasets that can be converted to having the form of a table (e.g. data held in key-value databases). It is possible that the techniques described in this document do not apply to more complex datasets, e.g. containing free-form text, images, audio, or video.

The use of de-identification techniques is good practice to mitigate re-identification risk, but does not always guarantee the desired result. This document establishes the notion of a formal privacy measurement model as an approach to the application of data de-identification techniques.

NOTE 1 Annex C clarifies how selected de-identification techniques described in this document are applicable for de-identification of free-form text.

NOTE 2 The application of de-identification techniques can be a privacy risk treatment option arising from a privacy impact assessment, as described in ISO/IEC 29134[32].

The selection of de-identification techniques needs to effectively address the risks of re-identification in a given operational context. There is therefore a need to classify known de-identification techniques using standardized terminology, and to describe their characteristics, including the underlying technologies and the applicability of each technique to the reduction of the risk of re-identification. This is the main goal of this document. The relationship between the terminology used in this document and related terminology used elsewhere (e.g. the notion of anonymization) is described in Annex B. However, the specification of detailed processes for the selection and configuration of de-identification techniques, including assessments of data usefulness and the overall risk from a re-identification attack, is outside the scope of this document.

NOTE 3 Authentication, credential provisioning, and identity proofing are also outside the scope of this document.

De-identification techniques are typically accompanied by technical and other organizational measures to enhance their effectiveness. The use of these measures is also described wherever applicable.

This document provides an overview of core concepts relating to the de-identification of data, and establishes a standard terminology for, and description of, the operation and properties of a range of de-identification techniques. However, it does not specify how these techniques should be managed in a particular use case. It is anticipated that sector-specific framework standards will be developed to provide such guidance.

Privacy enhancing data de-identification terminology and classification of techniques

1 Scope

This document provides a description of privacy-enhancing data de-identification techniques, to be used to describe and design de-identification measures in accordance with the privacy principles in ISO/IEC 29100.

In particular, this document specifies terminology, a classification of de-identification techniques according to their characteristics, and their applicability for reducing the risk of re-identification.

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*