ICS 35.030
CCS L80

NATIONAL STANDARD OF THE PEOPLE'S REPUBLIC OF CHINA

GB/T 42460-2023

Information security technology - Guide for evaluating the effectiveness of personal information de-identification

信息安全技术 个人信息去标识化效果评估指南

Issued on: March 17, 2023 **Implemented on: October 1, 2023**

Issued by: State Administration for Market Regulation;
Standardization Administration of the PRC

Table of Contents

Foreword

This document was issued on 17 March 2023 by the State Administration for Market Regulation; Standardization Administration of the PRC and takes effect on 1 October 2023.

It is a GB/T standard: recommended rather than compulsory, but it is the text a Chinese reviewer applies when assessing a submission.

1 Scope

GB/T 42460-2023 covers how far a set of personal data has actually been de-identified and how that is proved - a grading of identifiability into four levels, the process an evaluation follows, and the work itself: preparation, a qualitative pass, a quantitative pass, the forming of conclusions, the communication and negotiation that follows, and the documentation kept

of the whole process. Stripping names and identity numbers out of a dataset feels as though it settles the matter, and it does not: dates, locations, job titles and similar quasi-identifiers combine to single out individuals in data that looks anonymous, and a release judged safe on its own can be re-identified once someone joins it to something else. The annexes work on exactly that difficulty, with examples of direct identifiers, examples of quasi-identifiers, a method for deciding which fields in a real dataset behave as quasi-identifiers, and a worked evaluation using the K-anonymity model. Written for organisations that publish, share or sell datasets derived from personal information, for the teams inside them who have to sign a release off, and for assessors and regulators checking whether a de-identification claim holds up.

This document provides guidelines for grading and evaluating the effectiveness of personal information de-identification.

This document applies to personal information de-identification activities. It is also applicable to personal information security management, supervision and evaluation.

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.

GB/T 25069-2022, Information security techniques -- Terminology

GB/T 35273-2020, Information security technology -- Personal information security specification

GB/T 37964-2019, Information security technology -- Guide for de-identifying personal information

3 Terms and definitions

For the purposes of this document, the terms and definitions defined in GB/T 25069-2022, GB/T 35273-2020, GB/T 37964-2019 as well as the followings apply.

3.1 personal information

Various information related to identified or identifiable natural persons recorded electronically or otherwise.

NOTE: Anonymized information is not included.

[Source: GB/T 35273-2020, 3.1, modified]

3.2 personal information subject

The natural person identified or associated with the personal information.

3.9 completely public sharing

Once the data is released, it is difficult to recall, and it is generally released directly through the Internet.

[Source: GB/T 37964-2019, 3.12]

3.10 controlled public sharing

The use of data is constrained by the data use protocol.

[Source: GB/T 37964-2019, 3.13]

3.11 enclave public sharing

Share within physical or virtual jurisdictions. Data cannot be exported outside the territory.

[Source: GB/T 37964-2019, 3.14]

3.12 re-identification risk; identifiability

The probability that the subject of personal information can be identified from the data.

3.13 equivalence class

A collection of rows in microdata where all quasi-identifier attribute values have the same value.

3.14 acceptable risk threshold

The set re-identification risk threshold value.

NOTE: When the re-identification risk is greater than this value, mitigation measures (including de-identification processing) and emergency measures need to be taken to keep the risk within a controllable range.

4 Grading of personal information de-identification effectiveness

Based on whether the data can directly identify the subject of personal information, or how likely it is to identify the subject of personal information, the identifiability of

personal information is graded into four levels, see Table 1, used to distinguish the effectiveness of de-identification of personal information.

Table 1 -- 4 levels of personal information identifiability

Grading Grading basis

c) Form an evaluation team, including personal information protection compliance experts, de-identification technical experts, and relevant business experts.

d) Carry out preliminary research, including detailed research on the data usage environment.

e) Determine the evaluation basis, including relevant laws, regulations and standards.

f) Determine the re-identification risk calculation scheme and acceptable risk threshold:

1) The re-identification risk calculation scheme considers both the dataset and the context in which it is used. It can be based on K anonymous model or differential privacy model, etc.

2) The acceptable risk threshold meets the corresponding safety requirements and meets the application needs.

g) Develop an evaluation plan.

6.2 Qualitative evaluation

Qualitative evaluation includes:

a) Identify the identifier according to 5.3 in GB/T 37964-2019. Form a list of identifiers (including direct identifiers and quasi-identifiers).

b) Determine whether the dataset contains identifiers in the identifier list. If it does not contain any identifiers, it is rated as level 4 and the evaluation ends; otherwise continue.

c) Determine whether the dataset has eliminated direct identifiers from the identifier list. If it contains the direct identifiers in the list, it is rated as level 1, and the evaluation ends; otherwise, further quantitative evaluation is carried out.

6.3 Quantitative evaluation