DRAFT International Standard

ISO/IEC DIS 10267

Information technology — Data usage — Methods to quantify the amount of personal information in a dataset

ICS: 35.030

ISO/IEC JTC **1**/SC **27**

Secretariat: **DIN**

Voting begins on:
2026-06-10

Voting terminates on:
2026-09-02

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENTS AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

This document is circulated as received from the committee secretariat.

Reference number
ISO/IEC DIS 10267:2026(en)

© ISO/IEC 2026



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2026

All rights reserved. Unless otherwise specified, or required in the context of its implementation, 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
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 ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Abbreviated terms ... **2**
5 Introduction ... **2**
5.1 PII processing overview ... 2
5.2 PII lifecycle ... 3
5.3 The PII processing context ... 4
6 Personally identifiable information and risks ... **5**
6.1 Data sensitivity ... 5
6.1.1 Overview ... 5
6.1.2 Different types of Personally identifiable information (PII) ... 5
6.1.3 Data classification process ... 7
6.2 The personal information factor (PIF) ... 8
6.3 The personal information factor versus risk of re-identification ... 8
6.4 Quantifying Risk with QaR ... 9
6.5 When should the two measurement tools be used? ... 10
7 Understanding the PIF method ... **10**
7.1 Overview ... 10
7.2 An approach based on information gain ... 10
7.2.1 Principles ... 10
7.2.2 Quantifying information content ... 11
7.2.3 Spatial, Temporal and Relationship features ... 11
8 Understanding the QaR method ... **12**
9 Additional framing of PII sharing and PII lifecycle ... **12**
9.1 Temporal, spatial, personal features and relationship features ... 12
9.2 Post data usage ... 13
Annex A (normative) **PIF and QaR calculation methodologies** ... **14**
Annex B (informative) **PIF and QaR use cases** ... **21**
Bibliography ... **29**

© ISO/IEC 2026 – 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 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 JTC1, 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.

© ISO/IEC 2026 – All rights reserved

Introduction

Sharing of large volumes of often personal data between individuals and organizations, or between individuals and governments, can give insights about the individual person, their relationships, habits and preferences and can raise confidentiality risk and negative consequences to the individuals.

Linking different datasets, even if de-identified, creates a great amount of personal information in the joined data, sufficient to reasonably re-identify individuals represented in the datasets.

How this data is used and by whom, and for what purposes creates risks and concerns.

The challenge is the need to address the big issues around sharing of de-identified data, broadly articulated as the need to protect individual privacy, concerns about unintended consequences of data sharing and the sensitivity of the data which is to be shared.

The focus for this document is the characterisation of two measures:

— a measure for the amount of personal information in a linked de-identified dataset. This measure – a personal information Factor (PIF) – is a measure of uniqueness of the members of a dataset and the information gained from identifying these individuals;

— Quantile-at-Risk (QaR), which is a way to measure how likely it is that someone could be identified from a dataset. It looks at how certain combinations of variables—like age, location, or occupation—might make some records stand out from the rest. If only a few entries have a rare mix of these details, those records could be more easily linked back to individuals. QaR helps estimate how exposed those rare, high-risk cases are, giving an overall sense of how well the dataset protects privacy.

NOTE The two measurement methods are only applicable to the features of a dataset that are structured data (i.e. excluding those containing data such as images, videos or audios).

This document doesn't provide guidance or criteria to evaluate or compare the results of the measure applied (PIF or QaR), nor guidance on how to articulate them with data de-identification techniques (such as those described in ISO/IEC 20889:2018).

For data or technology governance processes, organizations can refer to those outlined in the following International Standards:

— ISO/IEC 38500 Information technology – Governance of IT for the organization

— ISO/IEC 38505-1 Information technology – Governance of IT – Governance of data – Part 1: Application of ISO/IEC 38500 to the governance of data

In addition, for the suitability of IT systems, security and storage conditions, organizations can refer to the following standards:

— ISO/IEC 27001 Information technology – Security techniques – Information security management systems – Requirements

— ISO/IEC 27701 Security techniques –Privacy information management systems – Requirements and guidance

— ISO/IEC 27040 Information technology – Security techniques – Storage security

— ISO/IEC 5207 Information technology - Data Usage – Terminology and Use Cases

— ISO/IEC 5212 Information technology - Data Usage – Guidance for Data Usage.

© ISO/IEC 2026 – All rights reserved

Information technology — Data usage — Methods to quantify the amount of personal information in a dataset

1 Scope

The purpose of this document is to provide guidance on the methods to quantify the amount of personal information and to help estimate how easy it might be to re-identify someone in a dataset subjected to de-identification when it contains information about the characteristics of, actions of, or relationships to, natural persons. This guidance can further help organizations make better decisions for privacy protection and for appropriate use, sharing and exchange of such data.

This document is a high level, principles-based advisory standard which sets out terms and concepts that can be referenced by organizations.

This document does not provide an estimate of how easy it is to identify someone where additional external data is accessible, nor does it provide a way to define whether data is PII or not.

2 Normative references

ISO/IEC 5207:2024, *Information technology — Data usage — Terminology and use cases*

ISO/IEC 29100:2024, *Information technology — Security techniques — Privacy framework*

ISO/IEC 20889:2018, *Privacy enhancing data de-identification terminology and classification of techniques*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 5207:2024, ISO/IEC 29100:2024 and the following apply.

ISO and IEC maintain terminological 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
de-identification process
process of removing the association between a set of identifying attributes and the PII principal

[SOURCE: ISO/IEC 20889:2018, 3.6, modified – 'data principal' replaced by 'PII principal']

3.2
minimum identifiable cohort size
smallest group of individuals in a dataset which have the exact same values

Note 1 to entry: If there are unique individual(s) in a dataset, the minimum identifiable cohort size is 1.

3.3
personal information factor
measurement of the amount of personal information in a de-identified dataset

Note 1 to entry: The value of a PIF is between 0 and 1. A value of 1 indicates that an individual is identified.

© ISO/IEC 2026 – All rights reserved

3.4
Quantile-at-Risk
Measurement of the quantity of unusually specific records in a dataset which carry the highest risk of re-identification

3.5
quasi-identifier (size p)
set of p distinct fields (columns) of a dataset.

Note 1 to entry: the quasi-identifier of the columns with indices $j_1, j_2, \ldots, j_p$ where $1 \le j_1 < j_2 < \ldots < j_p \le p$ is denoted as $Q\left(j_1, j_2, \ldots, j_p\right)$

4 Abbreviated terms

CIG cell information gain

FIG feature information gain

GPD general pareto distribution

MICS minimum identifiable cohort size

PII personally identifiable information

PIF personal information factor

POT peak over threshold

QaR quantile-at-risk

RIG row information gain

5 Introduction

5.1 PII processing overview

PII processing is any activity involving PII. This includes the use, sharing and exchange of PII that can occur across entities of all types, sizes, and purposes. The decision-making process around PII usage requires the identification of:

— the decision to use, exchange or share PII;

— the purpose for PII use, exchange or sharing;

— PII characteristics, quality, security, and privacy;

— pathways to use, share and exchange PII and the alternatives;

— risks;

— mitigation measures for risks;

— authorization steps required;

— the policies, processes, procedures, or instruments required to provide predictability and reliability around PII processing.

© ISO/IEC 2026 – All rights reserved