ICS 35.030

中華民國國家標準

C N S

資訊安全、網宇安全及隱私保護－隱私增強資料去識別化框架

Information security, cybersecurity and privacy protection − Privacy enhancing data de-identification framework

CNS 27559(草-制1150246):2026

中華民國　年　月　日制定公布
Date of Promulgation:　-　-

中華民國　年　月　日修訂公布
Date of Amendment:　-　-

目錄

前言

本標準係依據 2022 年 11 月發行之第 3 版 ISO/IEC 27559，不變更技術內容，制定成為中華民國國家標準者。

本標準係依標準法之規定，經國家標準審查委員會審定，由主管機關公布之中華民國國家標準。

依標準法第四條之規定，國家標準採自願性方式實施。但經各該目的事業主管機關引用全部或部分內容為法規者，從其規定。

本標準並未建議所有安全事項，使用本標準前應適當建立相關維護安全與健康作業，並且遵守相關法規之規定。

本標準之部分內容，可能涉及專利權、商標權與著作權，主管機關及標準專責機關不負責任何或所有此類專利權、商標權與著作權之鑑別。

簡介

Introduction

去識別化係一種潛在手段，其以不識別之方式，或以不以其他方式危害個人隱私或個人群組隱私的方式，促進個人可識別資訊(PII)之使用。適切使用去識別化技術，可支援對法規要求事項及相關隱私原則之遵循性。然而，本標準中使用之用語〝資料當事人〞較〝PII 當事人〞更廣義，例：〝資料當事人〞包括組織及電腦。

De-identification is one potential means for facilitating the use of personally identifiable information (PII) in a way that does not identify or otherwise compromise the privacy of an individual or a group of individuals. The appropriate use of de-identification techniques can support compliance with regulatory requirements and relevant privacy principles. However, the term “data principal〞 used in this document is broader than “PII principal〞 and, for example, includes organizations and computers.

於幾乎所有情況下，去識別化至少要求評估個人或群組可獲取之額外資訊，此等資訊可能不適切揭示或暴露 PII (無論資料當事人是否遭蓄意識別，此等個人或群組皆稱為對手)，並要求評估對手如何將此等資訊結合，從而揭示或暴露 PII。簡而言之，去識別化要求評估資料提供予資料接收者之環境及情況。此考量對手可獲取何等額外資訊、攻擊之可能性及重新識別的動機。去識別化亦要求對資料進行評估。此判定對手如何能利用可獲得之額外資訊，揭示或暴露 PII，以及藉由自身或推論攻擊，進行重新識別或身分揭露的可能性。

In almost all cases de-identification requires, at the very least, an evaluation of the additional information available to an individual or group that can inappropriately reveal or uncover PII (which is referred to as an adversary, whether a data principal is identified intentionally or not), and how they can combine it to reveal or uncover PII. In short, de-identification requires an assessment of the environment and the circumstances in which the data are made available to data recipients. This considers what additional information is available to an adversary and the possibility of attacks and motivation to re-identify. De-identification also requires an assessment of the data. This determines how the additional information available to an adversary can be used to reveal or uncover PII and the possibility of re-identification, or identity disclosure, by itself or attacks of inference.

本標準向組織提供實作框架，用以治理 CNS 20889 中所述之資料去識別化技術的適切使用。此去識別化框架，可套用於資料生命週期之所有階段：從設計資料蒐集方法、該資料的內部重新使用、向外部合作夥伴提供資料，或歸檔。因此，資料接收者可能係資料保管者之內部人員，或外部人員，資料保管者依此去識別化框架，實作程序及實務作法。如圖 1(a)所示，使用及重新使用意味著保管者於向內部部門或職能群組，提供經去識別化資料之同時，仍對其維持監督。圖 1(b)顯示外部分享，其意味著保管者對經去識別化資料維持監督，同時將其提供予外部資料接收者(例：透過虛擬存取入口網站或實體資料中心)。圖 1(c)顯示外部發布，其意味著保管者將經去識別化資料之監督，移交予外部資料接收者。於此等情況之各情況下，去識別

化過程本身，可轉移予第三方，與保管者或接收者分開。與接收者之書面協議，依適用法律，判定去識別化後能使用所提供資料的方式。

This document provides organizations with an implementation framework to govern the appropriate use of data de-identification techniques described in ISO/IEC 20889. This de-identification framework can be applied at any point in the data lifecycle: from designing the means of data collection, the internal reuse of that data, making data available to external partners, or archival. The data recipients can therefore be internal or external to the data custodian that is implementing procedures and practices in accordance with this de-identification framework. As shown in Figure 1 a), use and reuse implies the custodian maintains oversight over the de-identified data while making it available to an internal department or functional group. Figure 1 b) shows external sharing, which implies the custodian maintains oversight over the de-identified data while making it available to an external data recipient (e.g. through a virtual access portal, or a physical data centre). Figure 1 c) shows external release, which implies the custodian transfers oversight over the de-identified data to an external data recipient. In each of these cases, the process of de-identification itself can be transferred to a third party, separate from the custodian or recipient. Written agreements with the recipient determine how data made available after de-identification can be used, in accordance with applicable laws.

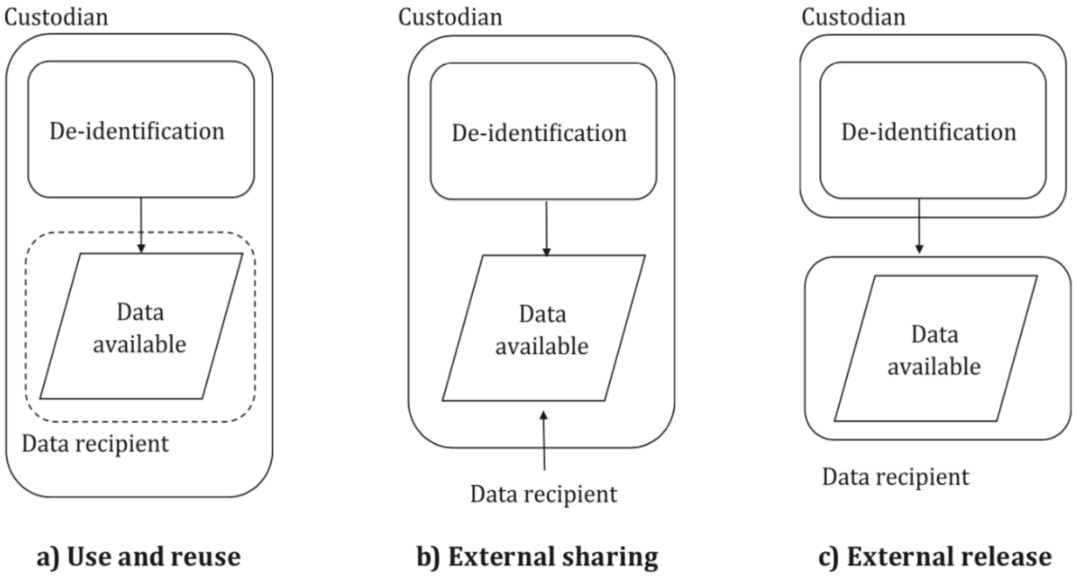

圖 1　資料可用性

Figure 1 − Data availability

圖 1 中之英中對照

Custodian	保管者
De-identification	去識別化
Data available	資料提供
Data recipient	資料接收者
a) Use and reuse	(a) 使用及重新使用
b) External sharing	(b) 外部分享
c) External release	(c) 外部發布

1. 適用範圍

1 Scope

本標準提供框架，用於識別及緩解重新識別風險，以及與經去識別化資料生命週期相關聯之風險。

This document provides a framework for identifying and mitigating re-identification risks and risks associated with the lifecycle of de-identified data.

本標準適用於所有型式與規模之組織，包括公開發行公司及私有公司、政府機關(構)與非營利組織，其係代表控制者的 PII 控制者或 PII 處理者，針對隱私之增強目的，實作資料去識別化過程。

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. 引用標準

2 Normative references

下列標準因本標準所引用，成為本標準之一部分。下列引用標準適用最新版(包括補充增修)。

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.

CNS 20889　隱私增強資料去識別化術語與技術分類

CNS 27000　資訊技術－安全技術－資訊安全管理系統－概觀及詞彙

CNS 29100　資訊技術－安全技術－隱私框架

CNS 31000　風險管理－指導綱要

ISO/IEC 27000, Information technology − Security techniques − Information security management systems − Overview and vocabulary

ISO/IEC 29100, Information technology − Security techniques − Privacy framework

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

ISO 31000, Risk Management − Guidelines

3. 用語及定義

3 Terms and definitions

CNS 20889、CNS 27000、CNS 29100、CNS 31000 所規定及下列用語及定義適用於本標準。

For the purposes of this document, the terms and definitions given in ISO/IEC 27000, ISO/IEC 29100, ISO/IEC 20889, ISO 31000 and the following apply.

~~ISO 及 IEC 於下列網址維護用於標準化之辭彙資料庫：~~

~~－ ISO 線上瀏覽平台：可於 https://www.iso.org/obp 取得。~~

~~— IEC Electropedia：可於 http://www.electropedia.org/取得。~~

ISO and IEC maintain terminology 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 保管者(custodian)

3.1 custodian

個人或個體，對電子儲存資訊，具保管、控制或持有權。

[來源：ISO/IEC 27050-1:2019 之 3.2]

person or entity that has custody, control or possession of electronically stored information

[SOURCE: ISO/IEC 27050-1:2019, 3.2]

3.2 資料接收者(data recipient)

3.2 data recipient

個人或組織，資料由其存取、與其分享或對其發布。

person or organization by, with or to whom data is accessed, shared or released

3.3 對手(adversary)

3.3 adversary

個人或單位，其能(無論蓄意或非蓄意)利用潛在脆弱性。

備考：於去識別化文獻中，對手、攻擊者、入侵者、窺探者及其他類似用語，經常可互換使用。

individual or unit that can, whether intentionally or not, exploit potential vulnerabilities

Note 1 to entry: Adversary, attacker, intruder, snooper, and other similar terms are often used interchangeably in the de-identification literature.

3.4 威脅建模(threat modelling)

3.4 threat modelling

系統性探索技術，旨在揭示可能以破壞、*揭露*(3.8)、資料修改或阻絕服務等形式，對系統造成損害之所有情況或事件。

[來源：修訂自 ISO/IEC/IEEE 24765:2017 之 3.4290。]

systematic exploration technique to expose any circumstance or event having the potential to cause harm to a system in the form of destruction, *disclosure* (3.8), modification of data, or denial of service

[SOURCE: ISO/IEC/IEEE 24765:2017, 3.4290, modified − Note 1 to entry has been deleted.]

3.5 隱私衝擊評鑑(privacy impact assessment, PIA)

3.5 privacy impact assessment

PIA

關於處理個人可識別資訊(PII)，識別、分析、評估、諮詢、溝通及規劃可能隱私

衝擊之整體過程，其為組織更廣泛的風險管理框架之一部分。

[來源：修訂自 CNS 29134 之 3.7。]

overall process of identifying, analysing, evaluating, consulting, communicating and planning the treatment of potential privacy impacts with regard to the processing of personally identifiable information, framed within an organization's broader risk management framework

[SOURCE: ISO/IEC 29134:2017, 3.7, modified − Note 1 to entry has been deleted.]

3.6 經定義母體(defined population)

3.6 defined population

萃取自資料集之元素集，有助於*對手*(3.3)識別資料當事人的能力。

set of elements that a dataset is drawn from that contributes to the *adversary's* (3.3) ability to identify a data principal

3.7 樣本(sample)

3.7 sample

資料集，其僅係*經定義母體*(3.6)之一部分，因此，*對手*(3.3)無法確定任何特定個體是否包含於其中。

dataset that is only a proportion of the *defined population* (3.6), such that an *adversary* (3.3) cannot be certain that any particular entity was in it

3.8 揭露(disclosure)

3.8 disclosure

依已發現或遭利用之脆弱性，揭示資料集中的機密資訊或個人可識別資訊。

revealing confidential or personally identifiable information from a dataset based on a vulnerability that is found or exploited

3.9 經分享資料(shared data)

3.9 shared data

資料集，由保管者授予一組固定個體，對該等資料之存取權限。

dataset in which a fixed set of entities have been granted access to the data by the custodian

3.10 經發布資料(released data)

3.10 released data

資料集，其中保管者不再直接控制誰對該等資料具存取權限。

dataset in which the custodian no longer directly controls who has access to the data

3.11 資料隱私模型(data privacy model)

3.11 data privacy model

應用資料去識別化技術之作法，使可識別性的計算可行。

[來源：修改自 CNS 20889 之 3.3。]

approach to the application of data de-identification techniques that enables the calculation of identifiability

[SOURCE: ISO/IEC 20889:2018, 3.3, modified − The word "formal" and

“measurement″ have been deleted from the term and "data" added, and “re-identification risk“ has been replaced by “identifiability“ in the definition.]

3.12 書面協議(written agreement)

3.12 written agreement

資料分享協議、諒解備忘錄、資料存取請求、契約及所有其他正式登載之協議。

data sharing agreement, memorandum of understanding, data access request, contract and any other formally documented agreement

3.13 資料轉換(data transformation)

3.13 data transformation

資料之修改。

modification of the data

3.14 去識別化治理(de-identification governance)

3.14 de-identification governance

指導及控制去識別化過程之系統。

[來源：修改自 ISO/IEC 38500:2015 之 2.8。]

system of directing and controlling the de-identification process

[SOURCE: ISO/IEC 38500:2015, 2.8, modified − “the de-identification process″ has been added to the definition.]

4. 符號及縮寫

4 Symbols and abbreviated terms

PIA 隱私衝擊評鑑(privacy impact assessment)

PII 個人可識別資訊(personally identifiable information)

P 機率函數(probability function)

PIA privacy impact assessment

PII personally identifiable information

P probability function

5. 概觀

5 Overview

本標準之目標係提供原則式框架進行去識別化，該框架考量程序、風險及損害。原則式去識別化作法擬對實作及技術之具體細節保持中立。該框架分為 4 個主要部分：

The goal of this document is to provide a principles-based framework to approach de-identification, which considers procedures, risks, and harms. A principles-based approach to de-identification is intended to be neutral on the specifics of implementation and technologies. The framework is presented in four main parts:

− 全景(第 6 節)：對資料提供予資料接收者時之環境及情況的評估，以判定對手可能取得何等外部資訊。此意味著風險亦可透過全景控制措施(意指 IT 安全控制措施、書面協議中描述之義務，以及政策與治理措施)管理。

- 資料(第 7 節)：對資料之評估，以判定對手如何能利用可取得的額外資訊，揭示或暴露 PII。可藉由限制可取得之資料及資料的提供形式(藉由資料轉換)，管理風險。
- 可識別性(第 8 節)：評估可識別性之方法係全景風險(攻擊的機率)與資料風險(於遭受攻擊情況下，發生揭露之機率)的函數。應定義適切之許可差，以確保可識別性低於預先定義的許可差等級。
- 治理(第 9 節)：為保管者制定之登載的程序及實務作法，以確保現在及未來均能一致且有效完成上述工作，以及於提供經去識別化資料之前、期間及之後所要求的準備工作。
- Context (Clause 6): An assessment of the environment and circumstances in which the data are made available to data recipients, to determine what external information can be available to an adversary. This implies that risk can be managed through contextual controls as well (meaning the IT security controls, obligations described in written agreements, and policy and governance measures).
- Data (Clause 7): An assessment of the data, to determine how the additional information available to an adversary can be used to reveal or uncover PII. Risk can be managed by limiting what data are made available, and in what form that data will be made available (by transforming the data).
- Identifiability (Clause 8): The method of assessing identifiability is a function of context risk (the probability of an attack) and data risk (the probability of disclosure given that there is an attack). An appropriate tolerance shall be defined to ensure the identifiability is below a pre-defined tolerance level.
- Governance (Clause 9): Documented procedures and practices for the custodian to ensure the above are done consistently and effectively, now and in the future, and the preparations that are required before, during, and after de-identified data are made available.

若產生之經去識別化資料不符合(保管者或預期資料接收者)的驗收準則，則可能有必要重複此過程，以致力尋求雙方皆能接受之解決方案。例：可改善資料環境之隱私及安全實務作法，以致力降低威脅，並依資料接收者的運作全景，改善對其提供之資料的效用。整個作法可大致以線性方式理解，保管者之去識別化治理涵蓋全景評估、資料評估、可識別性評估與緩解措施，以及驗收準則等整體過程，如圖 2 所示。然而，可依實作需要(例：改善特定資料流之效率及可縮放性)，對各要素重新排序。CNS 31000 包含關於管理組織所面臨風險之指導綱要。

It can be necessary to repeat the process if the resulting de-identified data does not meet acceptance criteria (by the custodian or intended data recipient), in an effort to find solutions that are acceptable to both parties. For example, the privacy and security practices for the data environment can be improved in an effort to reduce threats and improve the utility of data that are made available to the data recipients

based on their operational context. The entire approach can be thought of in a somewhat linear fashion, with de-identification governance by the custodian embodying the overall processes of context assessment, data assessment, identifiability assessment and mitigation, and acceptance criteria, as shown in Figure 2. It is, however, possible to reorder elements based on implementation needs (e.g. improving efficiency and scalability for specific data flows). ISO 31000 contains guidelines on managing risk faced by organizations.

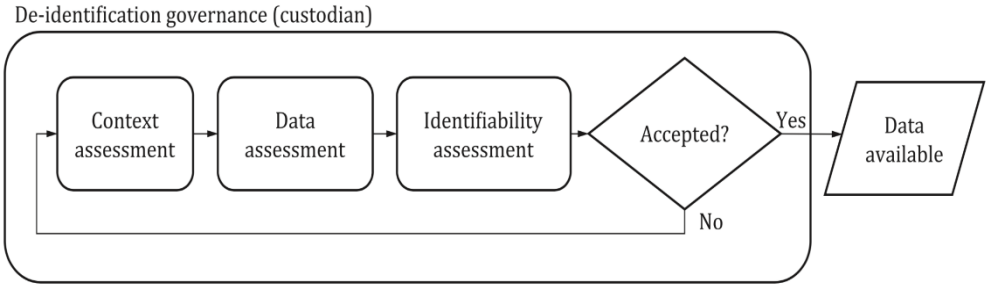

圖 2　實務作法中之去識別化框架

Figure 2 − De-identification framework in practice

圖 2 中之英中對照

De-identification governance (custodian)	去識別化治理(保管者)
Context assessment	全景評估
Data assessment	資料評估
Identifiability assessment	可識別性評估
Accepted?	接受？
Yes	是
No	否
Data available	資料提供

6. 全景評估

6 Context assessment

6.1 一般

6.1 General

保管者應評估向資料接收者提供資料之全景(無論係藉由提供經分享存取權限或藉由提供資料集複製本)，以協助妥善界定去識別化過程的範圍。

備考：可能套用法律要求事項。

The custodian shall evaluate the context in which data are being made available to a data recipient (either by providing shared access or by giving a copy of the dataset),

to help properly scope the de-identification process.

NOTE Legal requirements can apply.

判定此全景涉及對以下方面進行詳細評估：資料遭存取、分享或發布之環境；資料的來源及其預期用途；以及向資料接收者提供資料之具體情況(諸如透明度等級)。此等要素應以某種形式納入評估風險之方法中。例：可使用詳細之查檢表，對資料環境中的風險進行分類，並將其納入標準風險矩陣中，用以比較攻擊之可能性與於給定全景下，識別資料當事人的影響。然而，宜注意，僅當攻擊成功時，方發生揭露。

Determining this context involves a detailed assessment of the environment in which the data are accessed, shared or released, where the data come from and their intended use, and the circumstances under which the data are made available to a data recipient (such as levels of transparency). These elements shall be factored, in one form or another, into the method of assessing risk. For example, detailed checklists can be used to categorize risk in the data environment, and be factored into a standard risk matrix used to compare the possibility of an attack against the impact of identifying a data principal in a given context. It should be noted, however, that disclosures only occur if the attack is successful.

6.2 威脅建模

6.2 Threat modelling

6.2.1 一般

6.2.1 General

進行分享或發布之保管者，或進行評估的第三方，應使用客觀及結構化過程，評估資料將遭存取、分享或發布之環境。此環境包括人員及其動機(組織及個人)、其具權限存取之其他資料，以及其基礎設施與治理結構(包括 IT 安全控制措施，諸如 CNS 27002 及 CNS 27701 中所述的控制措施、存取政策等)。即使是自我管理之 IT 稽核或評估，亦能擷取大量關於發布環境的資訊，從而有助於界定潛在風險(尤其是潛在威脅)。

The custodian doing the sharing or releasing, or a third party doing an assessment, shall use an objective and structured process to evaluate the environment in which the data will be accessed, shared or released. This environment includes persons and their motives (organizational and individual), other data they have access to, and their infrastructure and governance structures (including IT security controls such as those described in ISO/IEC 27002 and ISO/IEC 27701, access policies, etc.). An IT audit or assessment, even self-administered, can capture a great deal of information regarding the release environment, to help frame potential risks (in particular, potential threats).

應使用結構化作法(通常稱為威脅建模)，評估可能揭示或暴露 PII 之攻擊風險。此包括檢查可能有哪些其他外部資料來源可用，並概述潛在揭露之對象、原因及方式。潛在威脅可能包括：

A structured approach, often known as threat modelling, shall be used to assess the risk of an attack that would reveal or uncover PII. This includes examining what other external data sources can be available and sketching out the who, why and how of a potential disclosure. Potential threats can be:

- 蓄意：針對性試圖揭示或暴露資料中之 PII，此等資料係由內部人員提供予作為資料接收者的群組或組織。
- Deliberate: A targeted attempt to reveal or uncover PII in the data that are made available to them by an insider to the group or organization that is the data recipient.
- 非蓄意：揭露亦可能係無意，例：當資料接收者處理經分享或經發布資料時，辨識出資料當事人。
- Accidental: A disclosure can also be unintentional, for example a data principal being recognized while a data recipient is working with the shared or released data.
- 環境因素：若已備妥之所有控制措施，均未能防止資料揭露，則資料亦可能遺失或遭竊。
- Environmental: The data can also be lost or stolen in the case where all the controls put in place have failed to prevent a data disclosure.

6.2.2 安全及隱私實務作法

6.2.2 Security and privacy practices

資料接收者之安全及隱私實務作法，將影響資料接收者場域的不法員工，能重新識別經分享資料之可能性。若缺乏強而有力之緩解控制措施，則不法員工可能選擇不遵守契約。安全及隱私實務作法，亦能判定外部人員直接或藉由破解內部人員信符，取得經分享資料之存取權限的可能性。

The security and privacy practices of the data recipient will have an impact on the likelihood of a rogue employee at the data recipient's site being able to re-identify the shared data. A rogue employee can choose not to abide by a contract in the absence of strong mitigating controls. The security and privacy practices can also determine the likelihood of an outsider gaining access to the shared data, either directly or by compromising an insider's credentials.

對緩解控制措施之評估，應詳盡且以證據為依據。

An evaluation of mitigating controls shall be detailed and evidence based.

備考：可能套用專業、國際及政府法規、標準及政策，包括 CNS 27002 及 CNS 27701 (適切時)。

NOTE Professional, international, and government regulations, standards, and policies can apply, including ISO/IEC 27002 and ISO/IEC 27701, where appropriate.

使用標準化作法，不僅確保單一組織分享資料之一致性，且確保跨組織的一致

性。

Using a standardized approach also ensures consistency, not only for a single organization that is sharing data, but across organizations.

為避免對資料提供者或接收者，造成不適切或過度之負擔，評估可將第三方稽核、相關驗證，以及重新使用先前分析，納入考量。

In order to avoid inappropriate or excessive burdens on the data provider or recipient, the evaluation can take into account third party audits and relevant certifications, as well as re-using prior analyses.

6.2.3 重新識別之動機及容量

6.2.3 Motives and capacity to re-identify

接收者揭示或暴露 PII 之動機，部分可藉由教育訓練、認知及書面協議中描述的義務(包括 ISO/IEC 23751 中所述之過程及條款)，加以控制，前提是其具強制力(透過法律機制及藉由拒絕分享或發布額外資料)。書面協議中之義務可包括：

A recipient's motives to reveal or uncover PII can be controlled in part by training, awareness, and obligations described in written agreements, including processes and terms described in ISO/IEC 23751, provided they are enforceable (through legal mechanisms and by refusing to share or release additional data). Obligations in written agreements can include:

- 向具權限存取經去識別化資料之個人，提供關於揭露風險的教育訓練。
- 定期提醒其履行維護資料隱私及安全政策之義務。
- 未經明確許可，禁止試圖於提供予資料接收者之資料中，識別當事人，或鏈結資料以延伸資料當事人的剖繪(從而增加揭露 PII 之風險)。
- 容許進行抽查或全面稽核(可能由第三方進行)，以確保遵循協議中規定之條款。
- 未經明確許可，禁止與任何其他資料接收者分享。
- 定義存取、分享及發布資料之環境(預期備妥基礎設施及治理結構)。
- 定義可接受之使用案例，以及對資料使用方式的所有預期，或如何評估使用案例，以確保其用於視為適切之目的。
- delivery of training on disclosure risks to individuals with access to de-identified data;
- regular reminders of their obligations to uphold data privacy and security policies;
- prohibiting attempts to identify data principals in the data made available to data recipients, or linking data that would extend the profiles of data principals (thereby increasing the risk of disclosing PII) without express permission;
- allowing for spot checks or full audits (possibly by a third party) that ensure compliance with the stated terms of the agreement;
- prohibiting the sharing with any other data recipient without express

permission;

- defining the environment in which the data are accessed, shared and released (the infrastructure and governance structures that are expected to be in place);
- defining acceptable use cases and any expectations of how the data are used, or how use cases are evaluated to ensure its use is for purposes that are deemed appropriate.

備考：當作為開放資料發布時，協議義務之執行可能不如預期有效。

NOTE It is possible that the enforcement of agreement obligations is not as effective when publishing as open data.

於規劃上述義務時，保管者宜瞭解資料之來源、去向，以及所存取、分享或發布資料的預期用途。此將有助於釐清資料發布之具體情況，以及向資料接收者提供資料的環境，從而採取適切措施，降低 PII 遭潛在揭露之風險。

In planning the above obligations, the custodian should know where the data have come from, where they are going, and what the intended uses are for the data being accessed, shared or released. This will help situate the data release and the environment in which they will be made available to data recipients, so that appropriate measures can be taken to reduce the risk of potential disclosures of PII.

6.3 透明度及影響評估

6.3 Transparency and impact assessment

6.3.1 一般

6.3.1 General

保管者應評估揭露之影響，該影響決定於評估可識別性時的許可差，且可包括：

The custodian shall assess the impact of disclosure, which decides tolerance when identifiability is evaluated, and can include:

- 盡可能保持透明，並於可行情況下，與利害相關者溝通。如此，相關個體能更理解對隱私之期望。
- 資料是否高度敏感且私密，是否來自弱勢族群，或是否可能揭示敏感屬性或具污名化性質。
- 因不適切處理而可能對資料當事人造成之潛在傷害或損害。
- 資料接收者之可信任度(例：備妥資料分享協議、合作歷程、維持持續合作夥伴的誘因)。
- 分享之目的及預期用途，如何與資料當事人的利益相符。
- 資料接收者將徹底考量，其對資料之潛在用途(例：隱私衝擊評鑑的使用)，並採取相對應行動之程度。
- 所有可能違反法律原則或基本權利之行為。
- Being as transparent as possible and engaging with stakeholders where practicable. This way expectations of privacy are better understood by the entities involved.

- Whether the data are highly sensitive and intimate, come from vulnerable populations, or can reveal sensitive attributes or be stigmatizing.
- Potential injury or harm to data principals that can arise due to inappropriate processing.
- The trustworthiness of the data recipients (e.g. data sharing agreement in place, history of partnership, incentives to remain an ongoing partner).
- How the purposes for sharing and intended uses are in line with the interest of data principals.
- The extent to which data recipients will thoroughly consider their potential uses of data (for example, the use of privacy impact assessments) and act accordingly.
- Any potential breach of legal principles or fundamental rights.

6.3.2 行動透明度及利害相關者參與

6.3.2 Transparency of actions and stakeholder engagement

保管者宜簡明扼要解釋其資料蒐集實務作法、其重新使用資料方式及其理由闡述，並對回饋保持開放態度，或就其資料分享活動，徵詢利害相關者之意見，以期瞭解並於適切時，解決其關切事項。信任要求開放。與利害相關者進行有意義之討論，可能有助於在風險與利益間，建立合理的平衡。

The custodian should explain simply and clearly its data collection practices, how it reuses data with a description of its rationale, and be open to feedback or seek the views of stakeholders on its data sharing activities with the goal of understanding, and where appropriate, addressing their concerns. Trust requires openness. Meaningful discussions with stakeholders can help to establish a reasonable balance of risk and benefits.

保管者可考量比較其行業內類似組織發布或分享資料之方式，以及關於其實務作法是否曾引發任何關切事項。此外，亦可考量進行調查及焦點小組研究，瞭解資料當事人對資料發布、分享及重新利用之看法(例：藉由產業協會)，以協助提供決策制定依據。

The custodian can consider comparing how similar organizations in its sector are releasing or sharing data, and whether any concerns have been raised about their practices. Surveys and focus group work on the data principals' views of data release, sharing and reuse (e.g., by industry associations) can also be considered to help inform decision making.

6.3.3 與隱私相關之損害

6.3.3 Privacy-related harms

保管者宜考量，因資料提供而可能造成之潛在隱私相關損害，包括因不適切處理而可能造成的傷害或損害。例：保管者可：

The custodian should consider the potential privacy-related harms that can result from data being made available, including potential injury or harm due to

inappropriate processing. For example, the custodian may:

(a) 評估資料是否高度敏感且涉及個人隱私。

(b) 考量可出現下列情況之使用案例：

(1) 揭示敏感屬性。

(2) 造成污名化。

(3) 支援對資料當事人產生不利影響之決策。

a) evaluate whether the data are highly sensitive and personal; and also

b) consider use cases that may:

1) reveal sensitive attributes,

2) be stigmatizing, or

3) support decisions that adversely affect data principals.

保管者宜進行隱私衝擊評鑑(PIA)，俾識別隱私相關風險及妥善控制措施，以緩解此等風險。PIA 過程可分享並解釋，以增加透明度。此過程始於倡議之最早期階段，此時仍有機會影響其結果，從而確保〝於設計即保護隱私〞。此過程將持續至專案部署完成，甚至部署之後。

A custodian should conduct privacy impact assessments (PIA) in order to identify privacy-related risks and the proper controls to mitigate them. The PIA process may be shared and explained to increase transparency. It is a process that begins at the earliest possible stages of an initiative, when there are still opportunities to influence its outcome and thereby ensure privacy by design. It is a process that continues until, and even after, the project has been deployed.

備考：CNS 29134 描述進行 PIA 過程之指導綱要，以及報告的結構與內容。

NOTE The guidelines for the process of conducting a PIA and the report's structure and content is described in ISO/IEC 29134.

宜考量建立監督機制，以避免或實質降低造成損害之風險，從而贏得資料當事人對資料使用的信任，且其中某些機制，宜作為義務轉移予資料接收者。

Oversight mechanisms should be considered that avoid or meaningfully reduce the risk of harm and thereby engender the data principal's trust in the use of data, and some of these should be transferred as obligations to data recipients.

7. 資料評估

7 Data assessment

7.1 一般

7.1 General

去識別化之目標係於保護資料當事人隱私的同時，產生符合使用案例要求事項之資料。資料評估係用以瞭解特徵，而攻擊建模則利用此等特徵評估潛在脆弱性(例：單獨挑出、鏈結及推論)，此等脆弱性描述於 CNS 20889 中。第 8 節中之可識別性評估依既定效標(benchmark)，定義臨限值或保護目標。預設之假設宜為，若資料的某個元素或特徵，對於該使用案例並非需要(敏感資訊、增加風險之複雜資料

結構等),則保管者宜規劃將其移除。保管者若能採取降低資料複雜度之任何措施,則轉而亦將降低其應進行的技術分析之複雜度。

The goal of de-identification is to produce data that meets the requirements of the use case, while protecting the privacy of data principals. A data assessment is used to understand features and attack modelling exploits these features to evaluate potential vulnerabilities (e.g. singling out, linking, and inference) which are described in ISO/IEC 20889. The identifiability assessment in Clause 8 defines a threshold or protection objective based on established benchmarks. The default assumption should be that if an element or feature of the data is not needed for the use case (sensitive information, complex data structures that increase risk, etc.), the custodian should plan to remove it. Anything that the custodian can do to reduce the complexity of the data will, in turn, reduce the complexity of the technical analysis that the custodian shall conduct.

7.2 資料特徵

7.2 Data features

7.2.1 一般

7.2.1 General

當考量是否及如何安全分享或發布資料時,關鍵考量因素在於資料本身。保管者應評估資料中屬可識別性核心之特徵。

When thinking about whether, and how, to share or release data safely, a key consideration is the data themselves. The custodian shall evaluate features of the data that are central to identifiability.

7.2.2 資料當事人

7.2.2 Data principals

保管者宜使用下列內容,以深入瞭解對手可取得之外部背景資訊:

The custodian should use the following to gain insight into the external background information available to an adversary:

- 資料集中所代表之資料當事人種類。
- 資料集中資料之本質及唯一性,以及其對對手的潛在價值。
- 對手可取得之其他關於資料當事人的資料。
- 具動機之對手透過針對性搜尋,能從中瞭解關於資料當事人的哪些資訊。
- the kind of data principals represented in the dataset;
- the nature and uniqueness of the data in the dataset and its potential value to an adversary;
- other data about a data principal available to an adversary;
- what a motivated adversary can learn about them through targeted searches.

保管者亦宜考量資料中是否具任何弱勢群組,並於決定是否需將此等群組納入其擬提供之資料中,以及決定保護此等群組所要求的過程時,予以特別關注。

The custodian should also consider whether there are any vulnerable groups in the

data and take extra care in deciding whether or not they are needed in the data the custodian wants to make available and the processes required to protect them.

7.2.3 資料型式

7.2.3 Data type

資料可能為：

- 單位層級紀錄，其中列代表單一母體單位(個人、家戶等)，行代表關於該等單位所蒐集之資訊(屬性)。
- 聚合資料，其中儲存格代表共享關於該等單位所蒐集之經分類資訊的母體單位之彙集。
- 非結構化資料，諸如內容或格式不受限制之自由形式文字，或上述的組合。

Data can be:

- Unit-level records, in which the rows represent a single population unit (person, household, etc.) and the columns represent the information (attributes) collected about them.
- Aggregated, in which cells represent the collection of population units that share categorized information collected about them.
- Unstructured, such as free-form text in which there are no restrictions on content or format, or a combination of the above.

備考：各種形式之資料皆承載不同的風險，因此需不同作法對其進行保護。

NOTE Each form of data carries different risks and different approaches are needed to protect them.

7.2.4 屬性型式

7.2.4 Attribute types

保管者應判定資料集之組成部分及其結構化內容，並於可行情況下，以紀錄層級精細度及欄位型式進行判定。例：查明資料元素是否可能為：

The custodian shall determine the constituent parts and structured content of the dataset, where feasible doing this at record level granularity and by field type. For instance, ascertaining whether data elements can be:

- 直接識別符。
- 間接識別符。
- 其他需關切之資訊(例：因資料的機密性或個人本質，而無法廣泛取得且易遭特定針對，可能需額外保護)。
- direct identifiers;
- indirect identifiers;
- of other concern (e.g. not widely available and targeted due to the confidential or personal nature of the data, possibly warranting additional protection).

附錄 A 提供直接識別符及間接識別符之示例。

Annex A provides examples of direct and indirect identifiers.

備考 1. 當處理敏感資料時，無論是蓄意試圖揭示或暴露 PII 之可能性，或是其造成的影響，其風險皆較高。資料越敏感，對組織於公眾信任及聲譽方面之影響可能越大。識別標的內容有助於組織瞭解揭露可能造成之損害。

NOTE 1 When dealing with sensitive data, the risk is higher both in terms of the likelihood of a deliberate attempt to reveal or uncover PII and the impact. The impact on the organization in terms of public trust and reputation is likely to be greater the more sensitive the data. Identifying target content informs the organization about possible harm that can arise from disclosure.

備考 2. 作為間接識別符之屬性，可能係資料接收者關注的屬性。此等屬性於可識別性評估中納入考量，且可將其移除。

NOTE 2 Attributes that are indirect identifiers can be the attributes of interest to the data recipients. They are considered in the identifiability assessment and can be removed

7.2.5 資料集性質

7.2.5 Dataset properties

保管者宜先考量資料集之性質如何可能增加或降低可識別性，以此作為進行仔細資料分析要求的前提。其考量因素可能包括：

The custodian should consider how properties of a dataset can potentially increase or decrease identifiability as a precursor to the requirement of careful data analysis and can include:

- 資料品質：所有資料皆包含某種程度之錯誤，此提供某種程度的保護。儘管如此，為盡可能提供最有用之資料，保管者可能仍希望最小化此錯誤。
  備考：關於資料品質模型之更多資訊，參照 ISO/IEC 25012。
- Data quality: All data contains some level of error which offers some degree of protection. Although the custodian is likely to want to minimize this error for the sake of providing the most useful data possible.
  NOTE Further information on data quality models is provided in ISO/IEC 25012.
- 群組結構資料：此係包含群組中彼此間互相鏈結之成員資訊的資料。該等資料提供(更多)資訊，可使資料當事人於經定義母體中具唯一性。
- 縱向資料：此係關於經定義母體，隨時間推移蒐集並鏈結之資料。可擷取到資訊隨時間推移發生之潛在唯一變更，並於其他縱向型樣中脫穎而出。
- 母體資料或樣本資料：母體資料係關於特定可識別群組中之所有人(相較於隨機樣本而言)。母體資料所代表之對象幾無不確定性。
- Group structure data: This is data that contains information for members of a group who are linked with one another. The data provide (more) information that can make a data principal unique in a defined population.
- Longitudinal data: This is data about a defined population which is collected

over time and linked. Potentially unique changes in information over time can be captured and stand out among other longitudinal patterns.

- Population or sample data: Population data are about all persons in a particular identifiable group (as opposed to a random sample). There will be little uncertainty as to who is represented in population data.

7.3 攻擊建模

7.3 Attack modelling

7.3.1 一般

7.3.1 General

保管者應識別一組潛在攻擊及其相關聯指標，以結合威脅建模(參照 6.2)，用於攻擊建模。

The custodian shall identify a set of potential attacks and associated metrics, which are used for the attack modelling, in conjunction with the threat modelling (see 6.2).

保管者可客觀考量下列攻擊，以判定其是否具可信賴之可能性，連同考量其他顯著的領域特定或資料集特定之關切事項：

The custodian may objectively consider the following attacks to determine if they are credible possibilities, along with other significant domain-specific or dataset-specific concerns:

- 對手知悉標的個別個體存在於資料中之攻擊。
- 對手不知或無法知悉標的個別個體是否存在於資料中之攻擊。
- 針對資料中可能存在之所有個體(而非標的個別個體)的攻擊。
- attack where the adversary knows that a target individual entity is in the data;
- attack where the adversary does not, or cannot, know if a target individual entity is in the data;
- attack on all entities (rather than a target individual entity) that can be in the data.

備考：於學術文獻中，此等風險有時分別稱為檢察官風險、新聞記者風險或行銷人員風險。

NOTE These are sometimes referred to as prosecutor, journalist, or marketer risk, respectively, in the academic literature.

指標亦可取所有資料當事人之平均值(平均風險)，或取所有資料當事人的最大值(最大風險)。

Metrics can also be averaged across all data principals (average risk), or the maximum taken across all data principals (maximum risk).

7.3.2 最大風險或平均風險

7.3.2 Maximum or average risk

- 最大風險：對於此指標，取單一資料當事人之可識別性的最高等級(對橫跨

經分享或經發布資料中之所有資料當事人，進行量測)。當未備妥任何控制措施(例：公開資料分享或發布)，以防止最具可識別性之資料當事人成為攻擊標的時，使用此指標。由於資料分享或發布之公開本質，保管者宜假設攻擊的可能性極高－即使僅為展現資料集中之潛在脆弱性，亦可能有人出於某種動機攻擊資料集(稱為展示攻擊)。

- Maximum risk: For this metric, the maximum level of identifiability for a single data principal is taken, measured across all data principals in the shared or released data. This metric is used when there are no controls in place (e.g. public data sharing or release) to prevent the most identifiable data principal from being targeted in an attack. The custodian should assume a high possibility of attack because of the public nature of the sharing or release of data − someone will likely be motivated to attack the dataset if only to demonstrate potential vulnerabilities in it (called a demonstration attack).
- 平均風險：對於此指標，取單一資料當事人之可識別性的平均等級(對橫跨經分享或經發布資料中之所有資料當事人，進行量測)。當未備妥任何控制措施(例：公開資料分享或發布)，以防止最具可識別性之資料當事人成為攻擊標的時，使用此指標，但未必僅限於最易識別的資料當事人。當備妥合適控制措施以防止展示攻擊時，保管者宜假設對手係〝誠實但好奇〞。
- Average risk: For this metric, the average level of identifiability for a single data principal is taken, measured across all data principals in the shared or released data. This metric is used when there are controls in place (e.g. non-public data sharing or release) to prevent any individual data principal from being targeted in an attack, but not necessarily limited to the most identifiable data principal. The custodian should assume an honest but curious adversary when suitable controls are in place to prevent a demonstration attack.

7.3.3 母體攻擊或樣本式攻擊

7.3.3 Population or sample-based attack

對手掌握關於其標的特定資料當事人之某些背景資訊，並利用此等背景資訊，於提供予接收者的經去識別化資料中，搜尋匹配之紀錄。攻擊分為 3 種型式，其中 2 種型式，可由保管者，依對手是否知悉標的資料當事人是否存在於所提供之資料中，進行評估，而第 3 種型式則指針對非特定標的之攻擊：

The adversary has some background information about a specific data principal they are targeting and uses this background information to search for a matching record in the de-identified data made available to recipients. There are three types of attacks, two of which the custodian may evaluate based on whether the adversary knows if a targeted data principal is in the data made available, and a third type when the attack is not targeted:

- 母體中個體之攻擊：於此情境下，對手知悉標的個體位於提供資料中經定義

母體內。依定義，標的個體即為資料當事人。

- 樣本中個體之攻擊：於此情境下，對手不知(或無法知悉)標的個體是否位於已提供資料中，極可能因標的個體位於較大之經定義母體的樣本中。標的個體可能為資料當事人，或非資料當事人。
- 資料集攻擊：於此情境下，對手欲於已提供之資料集中，揭露盡可能多的資料當事人之身分。
- Attack of entity in population: In this scenario, an adversary knows the targeted entity is in a defined population in the data made available. The targeted entity will, by definition, be a data principal.
- Attack of entity in sample: In this scenario, an adversary does not, or cannot, know if the targeted entity is in the data being made available, most likely because the targeted entity is in a sample of a larger defined population. The targeted entity can be a data principal, or not.
- Dataset attack: In this scenario, an adversary wants to disclose the identity of as many data principals as possible in a dataset that has been made available.

7.3.4 資料隱私模型

7.3.4 Data privacy models

保管者宜選擇資料隱私模型，藉由預測資料隱私度量(measure)，以協助量化資料中之可識別性。

備考：有關技術隱私指標(metric)分類法之說明，參照 D Wagner I 及 Eckhoff[21]。

The custodian should choose data privacy models to help quantify identifiability in the data by forecasting a measure of data privacy.

NOTE For a description of a taxonomy of technical privacy metrics, see D Wagner I and Eckhoff[21].

模型對對手及資料提供全景做出假設。如 CNS 20889 所述，此等模型有助於評估可於資料中直接或間接識別資料當事人之可能性，並有助於就降低資料可識別性所需的轉換(例：概化、抑制、聚合、雜訊添加)做出知情決策。然而，保管者於做出決策時，應考量資料存取、分享或發布之全景及資料隱私模型。

Models make assumptions about the adversary and the context in which the data are made available. They can be useful in assessing the likelihood that data principals can be identified directly or indirectly in the data and making informed decisions about what transformations to the data are needed to reduce identifiability (for example, generalization, suppression, aggregation, noise addition), as described in ISO/IEC 20889. However, the custodian shall consider the context of the data access, sharing or release as well as the data privacy model when making its decision.

有許多模型可供選擇。一類模型考量將個體鏈結至記錄、屬性或已提供之資料集本身。另一類模型考量分享或發布資料將獲得何等資訊。無論使用何種模型，

理解其假設及參數，對於判定需採取何等隱私措施以防止身分揭露，以及如何將資料發布之全景納入評估，至關重要。

There are many models to choose from. One class of models considers linking an entity to a record, to an attribute, or to the made available dataset itself. Another class of models considers what information will be gained in sharing or releasing data. Regardless of what model is used, understanding the assumptions and parameters is critical to the determination of what privacy measures are needed to prevent identity disclosure and how to factor the context of the data release into that assessment.

8. 可識別性評估及緩解措施

8.1 一般

8 Identifiability assessment and mitigation

8.1 General

保管者應將全景評估及資料評估納入重新識別風險之評估方法中，以判定資料的整體可識別性。可識別性係攻擊發生之機率(全景評估)與於發生攻擊的情況下，成功識別資料當事人之機率(資料評估)的函數。

The custodian shall factor the context assessment and data assessment into the method of assessing re-identification risks which determine the overall identifiability of the data. Identifiability is a function of the probability of an attack (context assessment) and the probability of successfully identifying a data principal given that there is an attack (data assessment).

8.2 評估可識別性

8.2.1 一般

8.2 Assessing identifiability

8.2.1 General

於資料分享或發布之全景下，保管者應考量各種揭露情境，以將其對可識別性的評估，建立於合理事件的框架之上。於建構此等情境時，保管者應考量潛在對手現實中可能存取之所有資料來源。若其他保管者已發布類似資料一段時間且未出現明顯問題，則保管者為此投入的資源可能相對有限。然而，保管者宜從諸如具數十年去識別化經驗之國家統計機構等信譽良好的組織，所進行的資料分享或發布中，尋求強而有力之先例。

In the context of data sharing or release, the custodian shall consider disclosure scenarios to ground the custodian’s assessment of identifiability in a framework of plausible events. In constructing these scenarios, the custodian shall consider all the sources of the data to which the would-be adversary can realistically have access. If other custodians have been releasing similar data for a while without any apparent problems, the resources that the custodian shall devote to this can be more modest. The custodian should look for strong precedents, however, from data sharing or releases by reputable organizations such as national statistical

organizations with decades of experience in de-identification.

保管者宜避免過度聚焦於資料中顯而易見之脆弱性。例：若於揭露情境中，無其他紀錄分享於所考量之識別屬性上的值組合，則該記錄即為唯一(稱為唯一性)。唯一性確實表明脆弱性，但若無可利用該唯一性之合理情境，則不可能發生身分揭露。保管者應透過模擬或建模，確認於情境中，所選定屬性與保管者資料集中之唯一記錄的匹配係屬正確。此作法考量到，於經定義母體中，該唯一記錄可能具統計孿生(statistical twin)，而該孿生記錄並未包含於情境樣本中。

The custodian should avoid focusing too closely on apparent vulnerabilities in the data. For example, a record is unique if no other records share its combination of values on the identifying attributes considered in a disclosure scenario (called uniqueness). Uniqueness does indicate vulnerability, but if there is no well-formed scenario through which that uniqueness can be exploited, no identity disclosure can happen. The custodian shall confirm that a match against a unique record in the custodian's dataset of the selected attributes in the scenario is correct, either through simulation or modelling. This takes into account that the unique record can have a statistical twin in the defined population that is not represented in the scenario sample.

備考：屬性揭露可能協助導致身分揭露，並可能納入資料隱私模型中，如 CNS 20889 所述。

NOTE Attribute disclosures can help enable identity disclosures and can be factored into data privacy models, as described in ISO/IEC 20889.

8.2.2 量化可識別性

8.2.2 Quantifying identifiability

保管者應使用客觀方法及公認之效標，以選擇可識別性臨限值。此臨限值應與所使用之資料隱私模型配合，並於給定運作全景下，掌握識別資料當事人的可能性及其影響。對於公開資料，此臨限值可依最大風險而定；而對於非公開資料，該臨限值則可依平均風險而定。

The custodian shall use objective methods and well-established benchmarks for selecting an identifiability threshold. The threshold shall work with the data privacy model that is being used and capture the likelihood and impact of identifying a data principal in a given operational context. For public data, the threshold can be based on maximum risk; for non-public data, the threshold can be based on average risk.

可識別性能依循標準風險模型進行量化，其中識別機率(風險發生之可能性)等於在給定威脅下識別的機率乘以威脅實現之機率。亦即，

Identifiability can be quantified following a standard risk model where the probability of identification (the likelihood of a risk) is given by the probability of identification given a threat times the probability of a threat being realized. That is,

P (識別) = P (識別|威脅) × P (威脅)。

備考：條件機率 P(A | B)係指於已知 B 之情況下，A 發生的機率，或於 B 發生之情況下，A 發生的機率。

P(identification) = P(identification | threat) × P(threat).

NOTE The conditional probability, P(A | B), is the probability of A given B, or the probability that A will occur on the condition that B occurs.

附錄 B 彙總公認之可識別性臨限值。

Annex B summarizes well-established identifiability thresholds.

保管者可使用主觀及客觀準則，以影響影響力之評估，此評估可依 6.3 所列項目。若影響力視為高，則決策宜較傾向於採用較低之臨限值。反之，若影響力視為低，則可接受較高之臨限值。

A subjective and objective criterion may be used by the custodian to influence the evaluation of impact, which may be based on the items listed in 6.3. If the impact is deemed to be high, that should skew the decision more toward a lower threshold. On the other hand, if the impact is deemed to be low, a higher threshold would be acceptable.

當評估可識別性時，可將下列要件納入考量：

- 分析經分享資料對資料當事人或社會帶來何等合法利益？
- 資料是否高度詳細、高度敏感且具個人本質？
- 不適切處理資料可能對資料當事人造成何等潛在傷害？
- 資料當事人核可揭露資料之適切性為何？

The following elements may be taken into account when assessing identifiability:

- What are the legitimate benefits to data principals or society from analysing the shared data?
- Are the data highly detailed, are they highly sensitive and personal in nature?
- What is the potential injury to data principals from an inappropriate processing of the data?
- What is the appropriateness of approval by data principals for disclosing the data?

臨限值亦可能依資料分析對個體之益處進行調整。

The threshold can also be adjusted based on the benefits to entities from analysis of the data.

例：對於政府部門在保管者內部重新使用資料，以改善對其公民及居民之服務，或對於贊助者在保管者外部重新使用醫學試驗資料，以開發或改善相關疾病的治療方法，可證實較高之臨限值係屬合理。

EXAMPLE A higher threshold, can be justified for the internal reuse, within the custodian, of data by a government department improving services for their citizens and residents, or for external reuse, outside the custodian, of medical trials data

by the sponsor for the purposes of developing or improving treatments of related diseases.

於示例中提及之 2 種情況下，資料當事人可能預期資料的相容或合法之重新使用。然而，重要的是切實展現此對個體具潛在益處，且各方之期望一致。否則，於符合適切法律依據之情況下，可使用假名化，同時維持資料的可識別本質，而非將移除資料之可識別本質的去識別化。

In both cases mentioned in the example, data principals are likely to expect compatible or legitimate reuses of data. It is, however, important to meaningfully demonstrate that there are potential benefits to entities, and that expectations are aligned. Otherwise, where suitable for the appropriate lawful basis, pseudonymization can be used while maintaining the identifiable nature of data instead of de-identification which removes the identifiable nature of data.

然而，於考量可能之重新識別時，對手並不知何等名稱匹配正確。當對手試圖查證已匹配之名稱時，可能有必要引入額外的誤差因子，以反映此不確定性。

The adversary, however, does not know which names are correctly matched when considering possible re-identifications. When the adversary attempts to verify the names that were matched, an additional error factor can be necessary to capture this uncertainty.

無論保管者使用何種可識別性措施，此等措施皆宜係公認且經測試。廣泛採用之標準化作法，係其假設及合理性依據，更有可能經專家嚴格審查。

Whatever identifiability measures the custodian uses, they should be well-established and tested. A standardized approach, adopted widely, is more likely to have assumptions and justifications that have been scrutinized by experts.

8.2.3 對抗性測試

8.2.3 Adversarial testing

作為對資料集可識別性進行實際評估之一部分，保管者宜使用對抗性測試或滲透測試，藉由使用〝友善〞對手模擬攻擊，驗核所作假設。此可能既係參考性，亦係良好之實務作法，但需技能、專業知識、時間及資源。然而，於高風險之資料分享或發布情境中(諸如開放資料)，此點尤其重要。

The custodian should use adversarial or penetration testing to validate assumptions made by simulating attacks using "friendly" adversaries, as part of a practical assessment of a dataset’s identifiability. This can be both informative and good practice, but takes skill and expertise as well as time and resources. It is especially important, however, in high risk data sharing or release scenarios such as open data.

具動機之對手可視為相對有能力、能存取外部資料資源(諸如網際網路及公開文件)，並積極願意透過查詢，以發掘資訊的人。於基本分析中，通常不假設其具專業知識或高級電腦技能，或不假設其將訴諸犯罪手段。當然，若理由充分，且於保管者本身之揭露情境中，如此做係屬合理，則保管者可使用另一套關於對手可能運用的知識、技能及資源型式之假設(參照參考資料[16]及[19])。

A motivated adversary can be characterized as someone who is relatively competent, who has access to external data resources such as the internet and public documents, and is actively willing to make enquiries to uncover information. In basic analyses, they are not typically assumed to have specialist knowledge or advanced computer skills, or to resort to criminality. The custodian can of course use a different set of assumptions about the type of knowledge skills and resources that an adversary may bring to bear if they are well justified and to do so makes sense within the custodian's own disclosure scenarios (see References [16] and [19]).

對抗性測試能更精確模擬動機明確之對手可能採取的行動，其明確考量資料及匹配中之錯誤，且依真實的資料蒐集及真實外部資料。然而，其與一特定之演練緊密相關，因此未必能代表所有可能發生的情況。於實務中，保管者宜將資料分析技術與對抗性測試結合，而非僅依賴其中任何一種。

Adversarial testing mimics more precisely what a motivated adversary can do, explicitly takes into account errors in data and matching, and is based on real data gathering and real external data. However, it is tied very tightly to one particular exercise and therefore doesn't necessarily represent all of the things that can happen. In practice, the custodian should combine data analytical techniques with adversarial testing rather than relying solely on either one.

8.3 緩解措施

8.3.1 一般

8.3 Mitigation

8.3.1 General

去識別化本質上涉及下列 1 或 2 項：(i)將提供資料之環境，或(ii)待提供的資料本身。若保管者之可識別性分析(包括全景及資料)建議資料接收者需較強之控制措施，則保管者應建議資料接收者重新組態設定資料環境或變更資料。此外，於決定採用何種作法時，宜注意，表面揭露(即看似發生揭露，即使實際上並非如此)，有時對保管者造成之損害可能與實際揭露同樣嚴重。一旦套用緩解措施，宜藉由與臨限值進行比較，以重新評估可識別性。

De-identification essentially attends to either or both of i) the environment in which the data will be made available, or ii) the data to be made available. If the custodian's identifiability analysis (context and data) suggests that the data recipient needs stronger controls, the custodian shall suggest to the data recipient to reconfigure the data environment, or change the data. Also, in deciding on what approaches to apply, note that apparent disclosures (in which there seems to be a disclosure, even if that's not the case) can sometimes be just as harmful to a custodian as actual disclosures. Once mitigation measures have been applied, identifiability should be re-evaluated by comparing with the threshold.

8.3.2 重新組態設定環境

8.3.2 Reconfiguring the environment

重新組態設定環境，本質上涉及控制誰具存取權限、如何存取資料及存取目的。例：

Reconfiguring the environment essentially involves controlling who has access, how they access the data and for what purposes. For example:

- 僅容許於保管者本身之安全環境(虛擬或實體資料隔離區)內存取。
- 規定資料所必要之安全等級。
- 規定所有分析輸出，於其發布前，均須經保管者核對並批准。
- 規定可存取資料之人員。
- 規定其他組織性之緩解措施，諸如所要求的教育訓練及契約義務。
- allowing access only within the custodian's own secure environment (virtual or physical data enclaves);
- specifying the requisite level of security for the data;
- specifying that all analytical outputs be checked and sanctioned by the custodian before they are published;
- specifying persons who may access the data;
- specifying other organizational mitigation measures such as required training and contractual obligations.

對環境實施控制措施或加強控制措施，往往對風險產生相當顯著之影響，例：通常排除特定形式的攻擊，因此，若資料係屬敏感，則此等控制措施當然值得考量。

Placing or tightening controls on the environment tends to have quite significant effects on the risk, often ruling out particular forms of attack, for example, and so if the data are sensitive these controls are certainly worth considering.

8.3.3 轉換資料

8.3.3 Transforming the data

通常，保管者從相當固定之存取、分享或發布環境的初步方案開始，並依自身之使用案例，對將提供予資料接收者的資料進行變更。保管者宜從屬性型式開始：

Usually the custodian starts from a fairly fixed proposal of what the access, sharing or release environment is, and shall work on changing the data to be made available to the data recipient with the custodian's use case in mind. The custodian should start with the attribute types:

- 直接識別符宜藉由抑制或替換為非依識別資訊之隨機值或假名，予以消除(且於高風險資料存取、共用或發布情境中不可逆，諸如公開資料發布)。
- 間接識別符宜於無需資訊時，藉由抑制消除，或若要求將揭露風險降低至適切等級(依保管者之情境)，則使用下述技術進行轉換。
- 敏感資訊宜於無需資訊時，藉由抑制消除，或若要求將傷害或損害風險降低至適切等級，則使用下述技術進行轉換。

- Direct identifiers should be eliminated by suppression, or by replacing them with random values or pseudonyms that are not based on the identifying information (and irreversible in high-risk data access, sharing or release scenarios, such as public data releases).
- Indirect identifiers should be eliminated by suppression if the information is not needed, or transformed using techniques described below if required to reduce the risk of disclosure to an appropriate level (based on the custodian's scenario).
- Sensitive information should be eliminated by suppression if the information is not needed, or transformed using techniques described below if required to reduce the risk of injury or harms to an appropriate level.

一般而言，保管者宜能透過 CNS 20889 中所述之概化、抑制及取樣，將可識別性降低至適切等級，因此等方法較易於理解，且若使用正確，則可對可識別性產生有意義影響。

In general, the custodian should be able to reduce identifiability to an appropriate level through generalization, suppression, and sampling as described in ISO/IEC 20889 because these methods are easier to understand and can have meaningful impact on identifiability if used correctly.

- 概化，藉由增加數值區間大小或合併資訊種類，以降低所提供資訊之詳細程度。
- 抑制，從提供予資料接收者之資料集中，排除某些屬性，從而完全消除該等資訊於考量之外。
- 取樣，從提供之資料集中移除個體，導致無法確定特定母體單位是否實際包含於資料中。
- Generalization reduces the level of detail of the information provided, by increasing the size of numerical intervals or merging categories of information.
- Suppression excludes certain attributes from the dataset made available to the data recipient, eliminating that information entirely from consideration.
- Sampling removes entities from the dataset being made available, which creates uncertainty that a particular population unit is actually represented in the data.

如 CNS 20889 所述，主要替代方案係各種形式之資料失真，即透過操縱資料，俾干擾揭露策略，從而使對手無法確定攻擊是否成功的技術。應用資料失真，可能以不可預測且不透明之方式，影響資料效用，並使保管者面臨是否分享或發布關於失真的資訊的困難問題。雖存在複雜作法(例：合成資料)，然非屬本標準範圍。

The main alternatives as described in ISO/IEC 20889 are various forms of data distortion, techniques that manipulate the data in order to foil disclosure strategies

so that an adversary cannot be certain that an attack was successful. Applying data distortion can affect data utility in an unpredictable and non-transparent manner and leaves the custodian with the difficult question about whether or not to share or release information about the distortion. Sophisticated approaches exist (e.g. synthetic data) but are outside the scope of this document.

8.3.4 重新評估

8.3.4 Re-evaluation

保管者可重新運行其可識別性量測，俾查看重新組態設定環境或變更資料，對可識別性之影響。許多資料隱私模型亦指導保管者選擇資料轉換方式，前提是保管者已將其資料分享或發布之全景納入考量。

The custodian can rerun their identifiability measurements in order to see what impact reconfiguring the environment or changing the data has on identifiability. Many data privacy models also guide the custodian through the selection of data transformations, provided the custodian has already taken into consideration the context of the custodian's data sharing or release.

9. 去識別化治理

9.1 一般

9 De-identification governance

9.1 General

去識別化治理包含 3 大支柱：人員、過程及技術。去識別化治理框架中之人員及過程均由所使用的技術支援及啟用。保管者應正式制定並定期審查與其資料處理活動相關之治理，以及其於資料安全、處理、管理、儲存、分享或發布方面的原則、政策及程序。此等措施判定於資料提供前後，如何管理使用者與資料之關係，以及於發生揭露時需採取的步驟。

De-identification governance comprises three pillars: persons, process and technology. Persons and processes of a de-identification governance framework are supported and enabled by the technology used. The custodian shall formalize and periodically review governance related to its data processing activities and its principles, policies, and procedures for data security, handling, management and storage, and sharing or release. These determine how users' relationships with the data are managed before and after data are made available and the steps needed in the event of a disclosure.

與資料分享相關之原則、政策及程序，宜與保管者較廣泛的資訊及資料安全實務作法相整合，並納入其中。

The principles, policies and procedures related to data sharing should be integrated with, and incorporated into, the custodian's wider information and data security practices.

9.2 資料提供前

9.2.1 一般

9.2 Before data are made available

9.2.1 General

於提供資料前，宜備妥相對應之組織結構，以確保未來風險維持在可忽略不計的程度。宜評估共享之必要性、相稱性及其法律依據。

Organizational structures should be in place before data are made available to ensure risks remain negligible going forward. The necessity and the proportionality of the sharing, and the legal premises to do so should be assessed.

9.2.2 指定角色及責任

9.2.2 Assigning roles and responsibilities

宜判定保管者內部各成員之角色及責任。

- 指派負責人，負責授權及監督揭露控制過程，並確保其具履行此責任必要之技能及知識。
- 確保所有相關員工皆接受過合適之教育訓練，並瞭解其於資料處理、管理、分享及發布方面的責任。教育訓練形式包括：
  - 關於保管者資料處理活動原則及程序之內部教育訓練。
  - 關於諸如揭露控制議題及技術、資料安全、資料保護法等核心要素之外部教育訓練。
  - 實施員工保密協議。

Individual roles and responsibilities within a custodian should be determined.

- Identify a person who is responsible for authorizing and overseeing the disclosure control process and ensure that they have the necessary skills and knowledge to do this.
- Ensure that all relevant staff are suitably trained and understand their responsibilities for data handling, management, sharing and releasing. This can take the form of:
  - in-house training on the principles and procedures of the custodian’s data processing activities;
  - external training on core factors such as disclosure control issues and techniques, data security, data protection law etc.;
  - implementing a staff non-disclosure agreement.

9.2.3 制定原則、政策及程序

9.2.3 Establishing principles, policies and procedures

保管者用於提供資料之內部結構，宜包括涵蓋下列的原則、政策及程序：

A custodian's internal structures for making data available should include principles, policies and procedures that cover:

- 整個去識別化過程。
- 監視每次資料發布所帶來之未來風險隱患。
- 涵蓋保管者所有與其資料保護政策及程序相關之運作活動，維護全面的紀

錄保存系統，以確保具清晰之稽核存底。

- 對保管者之所有資料產品或整個組織進行隱私衝擊評鑑(PIA)。PIA 可：
  - 協助保管者知悉並因應所有特定之隱私議題。
  - 確保保管者活動之透明度，
  - 提升對保管者所作所為之信任。
  - 協助保管者遵循相關隱私法規。
- 識別並處理揭露控制可能有問題之案例。保管者亦宜考量於(處理疑難案例之)過程中何時，保管者宜尋求諸如監理機關或去識別化專家等機關(構)的協助及建議。
- 處理非蓄意揭露。依保管者之特定需要，宜考量針對不同的潛在非蓄意揭露，制定個別之政策，或制定單一的政策。無論選擇何種方式，保管者皆宜考量非蓄意揭露可能發生之方式及相對應的應對措施。
- The whole de-identification process.
- Monitoring future risk implications for each data release.
- Maintaining a comprehensive record-keeping system across all the custodian's operational activities related to their data protection policies and procedures to ensure there is a clear audit trail.
- Undertaking a privacy impact assessment (PIA) for all the custodian's data products or across the organization as a whole. A PIA can:
  - help the custodian be aware of and address any particular privacy issues,
  - ensure the transparency of the custodian's activities,
  - promote trust in what the custodian does, and
  - help the custodian to comply with relevant privacy legislation.
- Identifying and dealing with cases where disclosure control can be problematic. The custodian should also consider at what point in the process (in dealing with a difficult case) the custodian should seek external help and advice from bodies such as a regulator or de-identification experts.
- Dealing with unintended disclosures. Depending on the custodian's particular needs, it should consider developing separate policies related to different potential unintended disclosures, or develop a single policy. Whichever is chosen, the custodian should consider how an unintended disclosure can occur and how it will respond.

9.2.4 識別及管理資料揭露

9.2.4 Identifying and managing a data disclosure

保管者宜瞭解何為資料揭露，並識別出降低此種事件發生可能性之方式。保管者宜：

A custodian should understand what a data disclosure is and identify ways to reduce the likelihood of such an occurrence. The custodian should:

- 定義資料揭露。區分全景與資料可支援較細緻評估是否要求進行事件調查，以及資料揭露之可能性(參照表 1)。
- 識別與保管者資料狀況相關之資料揭露型式。
- 識別可能導致揭露之因素，諸如將未加密的光碟帶出工作場所後遺失，或非蓄意將資料透過電子郵件傳送予錯誤的收件者。全盤思考各種可能之揭露情境，能非常有用於協助保管者識別揭露如何源自於其日常處理活動，以及亦可能發生的錯誤、程序違規或惡意行為。
- 制定措施以限制/避免可能導致/促成揭露之因素。
- 明定保管者將如何處理違反此等措施之情事。
- Define a data disclosure. Making a distinction between context and data can support a more nuanced evaluation of whether an incident investigation is required, and the possibility of a data disclosure (see Table 1).
- Identify the types of data disclosure relevant to the custodian's data situation.
- Identify those factors likely to lead to a disclosure, such as the loss of an unencrypted disc taken out of the workplace or the accidental emailing of data to the wrong person. Thinking through a range of possible disclosure scenarios can be very useful in helping the custodian identify how a disclosure can arise from their usual processing activities, as well as what errors, procedural violations or malicious intent can also occur.
- Establish measures to limit/avert those factors likely to lead to/facilitate a disclosure.
- Establish how the custodian will address violations of these measures.

表 1　經去識別化資料之事件及漏洞情境

	事件	漏洞
全景	評估全景漏洞之可能性	評估揭露之可能性
資料	評估資料揭露之可能性	評估揭露之影響

Table 1 − Incident and breach scenarios for de-identified data

	Incident	Breach
Context	Evaluate likelihood of context breach	Evaluate likelihood of disclosure
Data	Evaluate likelihood of data disclosure	Evaluate impact of disclosure

9.2.5 與利害相關者進行溝通

9.2.5 Communicating with stakeholders

保管者宜與利害相關者進行互動及溝通，以判定其希望瞭解何等關於保管者之

處理活動，並判定分享資訊的適切程度之構成要素為何。利害相關者可能希望瞭解保管者處理活動之〝內容〞，諸如處理何等資料或於何等環境中進行處理。其亦可能希望瞭解保管者處理活動之〝如何〞，諸如如何避免揭露或保管者如何判定環境係屬安全。此等解釋不宜揭示可能協助對手攻擊系統之安全相關資訊，且宜以不揭示商業秘密或商業上、技術上敏感細節的方式進行。

The custodian should engage and communicate with stakeholders to determine what they would like to know about the custodian's processing activities and to determine what constitutes an appropriate level of information to share. It is likely that stakeholders want to know the "what" of the custodian's processing activities, such as what data, or in which environments. They are also likely to want to know the "how" of the custodian's processing activities, such as how disclosures are avoided or how the custodian determines an environment to be safe. Such explanations should not expose security-related information that can assist an adversary in attacking the system and should be done a way that does not expose trade secrets, or commercially or technically sensitive details.

9.3 資料提供後

9.3.1 一般

9.3 After data are made available

9.3.1 General

雖保管者之資料於提供時，可視為安全，然於中期內，情況可能並非如此。

While the custodian's data can be considered safe at the time of its being made available, it is possible that this is not the case in the medium term.

9.3.2 監視資料環境

9.3.2 Monitoring the data environment

隨著 IT 能力之持續進步、支援對資料更高層次的存取權限與其分析容量，以及可用資料量之不斷增加，意味著保管者宜考量其提供資料所處的資料環境可能發生之變化。可採取之措施包括：

Continuing advancements in IT capabilities, supporting ever-greater access to data and capacity for their analysis, and an ever-increasing amount of available data, mean that the custodian should consider the potential for change in the data environment in which the custodian has made data available. Possible actions include:

- 維護保管者所提供之所有資料的清冊。
- 追蹤經分享資料之直接接收者。
- 將擬議之分享及發布活動與過往的分享及發布活動進行比較，以考量不同發布間可能存在之關聯，從而導致揭露。
- 持續關注資料環境之變化及其對所存取、分享或發布資料的影響。例：
  - 藉由例：閱讀技術期刊/部落格、收聽相關播客或參加相關活動等方式，

隨時瞭解可能影響保管者資料狀況之新技術及安全的發展動態。

- 藉由與諸如行業群組及相關資料保護或隱私機構等相關保管者互動，監視關於資料分享、發布及傳播之法律或指引的變更。
- 藉由例：審查網際網路上可用之資訊，並透過諸如公共清冊、本地社區紀錄、房產仲介清單、專業清冊、圖書館等更多傳統來源，持續追蹤目前及新的潛在識別屬性公共資料來源。

- 進行定期對抗性測試(參照 8.2.3)。
- Keeping a register of all the data the custodian has made available.
- Keeping track of direct recipients of shared data.
- Comparing proposed share and release activities to past shares and releases to take account of the possibility of linkage between releases leading to a disclosure.
- Maintaining awareness of changes in the data environment and how these can impact their accessed, shared or released data. For example:
  - keeping abreast of developments in new technologies and security that can affect the custodian’s data situation by, for example, reading technology journals/blogs, listening to relevant podcasts or attending relevant events;
  - monitoring changes in the law or guidance on data sharing, releasing and dissemination by engaging with relevant custodians such as industry groups and the relevant data protection or privacy authority; and
  - keeping track of current and new public data sources of potentially identifying attributes by, for example, reviewing the information available on the internet and through more traditional sources such as public registers, local community records, estate agents’ lists, professional registers, the library, etc.
- Performing periodic adversarial testing (see 8.2.3)

9.4 事件發生時之緩解措施

9.4 Mitigation in case of incident

為降低揭露發生時之影響，保管者應能對資料揭露做出回應，並付諸行動計畫，包括：

To reduce the impact if a disclosure were to occur, custodians shall be able to respond to a data disclosure and put their plans into action, including:

- 遏止資料外洩。
- 評估並處理所有持續存在之風險。
- 向主管機關通報資料外洩事件，並於要求時，向資料當事人報告。
- 檢討並吸取教訓。
- 適用時，實施緩解措施，以減輕對資料當事人造成之所有傷害或損害。
- containing a breach;

- assessing and dealing with any ongoing risk;
- reporting the breach to the competent authorities and, if required, to the data principals;
- reviewing and learning lessons;
- as applicable, implementing mitigation measures for any injury or harm to a data principal.

保管者應備妥能協助處理非蓄意揭露之機制。此包括備妥有效之治理政策及程序，本質上敘明誰做什麼、何時做及如何做，並整體上支援透明文化。

Custodians shall put in place mechanisms that can help deal with an unintended disclosure. This includes having effective governance policies and procedures in place which essentially identify who does what, when and how, and generally support a culture of transparency.

- 保管者應確保其具強健之稽核存底：此展現保管者已依循所有正確程序，並識別出於其處理活動中(若有)有必要於何處進行變更，以防止類似事件再次發生。
- 保管者應制定危機管理政策：此政策宜敘明關鍵角色及責任，並詳細制定行動計畫。
- The custodian shall ensure it has a robust audit trail: This demonstrates that the custodian has followed all correct procedures, and identifies where, if at all, in their processing activities it can be necessary to make changes to prevent a similar occurrence.
- The custodian shall have a crisis management policy: This policy should identify key roles and responsibilities and detail an action plan.

保管者應確保對其處理活動進行定期審查。

The custodian shall ensure it undertakes a periodic review of their processing activities.

附錄 A
(參考)
識別符示例
Annex A
(informative)
Example identifiers

直接識別符可能係於給定運作全景中，唯一識別資料當事人之所有識別號碼、特性或代碼。常見之直接識別符包括(但不限於)：

A direct identifier can be any identifying number, characteristic or code that uniquely identifies a data principal in a given operational context. Common direct identifiers include, but are not limited to:

(a) 姓名。

(b) 國民身分證統一編號。

(c) 護照號碼。

(d) 駕照號碼。

(e) 詳細地址。

(f) 電子郵件位址。

(g) 電話號碼。

(h) 傳真號碼。

(i) 銀行帳號。

(j) 車輛識別號碼及序號，包括車牌號碼。

(k) 社會安全號碼。

(l) 健保卡號碼。

(m) 病歷號。

(n) 裝置識別碼及序號。

(o) 生物特徵辨識碼，包括指紋及聲紋等。

(p) 完整臉部照片及任何其他可比較圖像對。

(q) 帳號、證書號碼或許可證號碼。

(r) 網際網路協定(IP)位址號碼。

(s) 網頁統一資源定位符(URL)。

a) Name

b) Civil identification number

c) Passport number

d) Driver's license number

e) Address details

f) Email address

g) Phone number

h) Fax number

i) Bank account

j) Vehicle identifiers and serial numbers, including license plate numbers

k) Social security number

l) Health card number

m) Medical record number

n) Device identifier and serial number

o) Biometric identification codes, including fingerprints and voice prints, etc.

p) Full face picture images and any other comparable pairs of images

q) Account number, certificate number or license number

r) Internet protocol (IP) address number

s) Web universal resource locators (URLs)

間接識別符可能係於運作全景中，單獨無法唯一識別資料當事人之任何屬性，但與其他資訊結合後，能唯一識別資料當事人。常見之間接識別符包括(但不限於)：

An indirect identifier can be any attribute that does not uniquely identify the data principal alone in the operational context, but in combination with other information can uniquely identify the data principal. Common indirect identifiers include, but are not limited to:

(a) 性別。

(b) 出生日期或年齡。

(c) 事件日期(例：入院、手術、出院、就診相關日期)。

(d) 地理範圍(例：郵遞區號、建物名稱、地區)。

(e) 族裔背景。

(f) 國籍、出生地。

(g) 語言。

(h) 原住民身分。

(i) 可視少數族裔身分。

(j) 職稱、工作單位、部門及其他職業資訊。

(k) 婚姻狀況。

(l) 教育程度。

(m) 受教育年限。

(n) 總收入。

(o) 宗教信仰。

(p) 家庭用電量。

a) Gender

b) Date of birth or age

c) Date of event (e.g. admission, surgery, discharge, visit-related date)

d) Geographic range (e.g. zip code, building name, region)

e) Ethnic origin

f) Nationality, place of origin
g) Language
h) Aboriginal identity
i) Visible minority status
j) Job title, work unit, department and other occupational information
k) Marital status
l) Education level
m) Years of schooling
n) Total revenue
o) Religious beliefs
p) Power consumption of a household

附錄 B
(參考)
臨限值可識別性效標示例
Annex B
(informative)
Example threshold identifiability benchmarks

B.1 可識別唯一項

B.1 Identifiable uniques

對於任何合理複雜度之情境，用以對特定威脅進行建模之間接識別符子集，皆將導致非 0 之可識別性度量值。此係由於估計之本質所致，因此，從科學角度來看，臨限值宜視為對不確定性之許可差。下列係於運作全景中，應用於可識別性度量值時，估計不確定性所建立之臨限值示例。

The subset of indirect identifiers that are used to model a specific threat, for scenarios of any reasonable complexity, result in a measure of identifiability that is non-zero. This is due to the nature of estimation, and thresholds should therefore be considered a tolerance on uncertainty from a scientific perspective. The following are examples of established thresholds in the estimation of uncertainty as applied to identifiability measures in an operational context.

表 B.1 彙總僅從樣本中估計出可識別唯一項之公認臨限值(參照參考資料[14]及[20])。此等特定臨限值不宜詮釋為母體中唯一項之比例。反之，此等臨限值代表於估計樣本中唯一項可識別性時，可接受之不確定程度(可將其詮釋為，於樣本中存在唯一匹配之情況下，從母體中找到正確匹配的機率)。若母體中存在真正之唯一項，則隨著取樣比例趨近於 1，風險值將趨近於 1。

Table B.1 summarizes well-established thresholds for identifiable uniques estimated from only a sample (see References [14] and [20]). These particular thresholds should not be interpreted as the proportion of uniques in the population. Rather, these thresholds represent an acceptable degree of uncertainty in the estimation of identifiability for uniques in the sample (they can be interpreted as the probability of a correct match from the population given a unique match in the sample). If there are true population uniques, risk goes to 1 as sampling fraction goes to 1.

表 B.1　唯一項效標臨限值(對樣本中個體之攻擊)

情境	全景(矩陣)	臨限值
公開	攻擊可能性高，影響低。	平均值 0.005
	攻擊可能性高，影響中。	平均值 0.001
	攻擊可能性高，影響高。	平均值 0.0005
非公開	攻擊可能性低至中，影響低至中。	平均值 0.1
	攻擊可能性中，影響中。	平均值 0.05
	攻擊可能性中至高，影響中至高。	平均值 0.005

Table B.1 − Benchmark thresholds for uniques (attack of entity in sample)

Scenario	Context (matrix)	Threshold
Public	High possibility of attack, low impact	Avg 0,005
	High possibility of attack, medium impact	Avg 0,001
	High possibility of attack, high impact	Avg 0,000 5
Non-public	Low-med possibility of attack, low-medium impact	Avg 0,1
	Medium possibility of attack, medium impact	Avg 0,05
	Medium-high possibility of attack, medium-high impact	Avg 0,005

B.2 可識別群組

B.2 Identifiable groups

下列可識別性種類係依資料集中具相同識別值集之群組(可識別群組)的大小。對於公開資料，機率係依任一可識別群組之最小大小。通常，可識別群組係落於 10 至 20 名資料當事人範圍內，此等資料當事人於揭露情境中，透過間接識別符進行匹配。若對手試圖將已知資料當事人之名稱，隨機指配予可識別群組，則對手對任一資料當事人識別正確的機率係該群組大小之倒數，依循最大可識別性的最佳實務作法，該機率範圍為 0.1 (1/10)至 0.05 (1/20)。對於非公開資料，該機率係所有可識別群組之平均值(平均可識別性)，並包括對全景(以及由此產生的威脅)之評估。表 B.2 彙總公認之可識別性臨限值(參照參考資料[15]、[17]及[18])。

The following categories of identifiability are based on the size of groups with the same set of identifying values in the dataset (an identifiable group). For public data, the probability is based on the minimum size of any one identifiable group. Typically, the identifiable group is in the range of 10-20 data principals that match on the indirect identifiers in a disclosure scenario. If an adversary attempts to randomly assign the names of known data principals to an identifiable group, the probability of the adversary being correct for any one data principal is the inverse of the group size, which gives the probability 0,1 (1/10) to 0,05 (1/20), following best practice for maximum identifiability. For non-public data, the probability is an average across

all identifiable groups (average identifiability) and includes an assessment of context (and thereby threats). Table B.2 summarizes well-established identifiability thresholds (see References [15], [17] and [18]).

表 B.2　可識別群組之效標臨限值(對母體或樣本中個體之攻擊)

情境	全景(矩陣)	臨限值
公開	攻擊可能性高，影響低。	最大值 0.1
	攻擊可能性高，影響中。	最大值 0.075
	攻擊可能性高，影響高。	最大值 0.05
非公開	攻擊可能性低至中，影響低至中。	平均值 0.1
	攻擊可能性中，影響中。	平均值 0.075
	攻擊可能性中至高，影響中至高。	平均值 0.05

Table B.2 − Benchmark thresholds for identifiable groups (attack of entity in population or sample)

Scenario	Content (Matrix)	Threshold
Public	High possibility of attack, low impact	Max 0,1
	High possibility of attack, medium impact	Max 0,075
	High possibility of attack, high impact	Max 0,05
Non-public	Low-med possibility of attack, low-medium impact	Avg 0,1
	Medium possibility of attack, medium impact	Avg 0,075
	Medium-high possibility of attack, medium-high impact	Avg 0,05

對於可識別群組，試圖查證匹配名稱之不確定性，產生額外的誤差因子，預計落於 0 至 1 之範圍內，甚至可能更低，依運作全景而定。例：對於包含 20 名資料當事人之可識別群組，對手能正確匹配身分的機率為 0.05 (1/20)，但查證要求估計之查證因子 0.1，此導致 0.005 的揭露風險(參照參考資料[13])。

For identifiable groups, the uncertainty of attempts to verify the names that were matched produces an additional error factor estimated to be in the range of 0,1, if not lower depending on the operational context. For example, for an identifiable group of 20 data principals, the probability the adversary can correctly match an identity is 0,05 (1/20), but to verify requires the estimated verification factor of 0,1, which results in a disclosure risk of 0,005 (see Reference [13]).

參考資料

[1] ISO/IEC 23751, Information technology − Cloud computing and distributed platforms − Data sharing agreement (DSA) framework

[2] ISO/IEC/IEEE 24765:2017, Systems and software engineering − Vocabulary

[3] ISO/IEC 25012:2008, Software engineering − Software product Quality Requirements and Evaluation (SQuaRE) − Data quality model

[4] CNS 27002 資訊安全、網宇安全及隱私保護－資訊安全控制措施

[5] ISO/IEC 27050-1:2019, Information technology − Electronic discovery − Part 1: Overview and concepts

[6] CNS 27701 安全技術－用於隱私資訊管理之 CNS 27001 及 CNS 27002 延伸－要求事項及指導綱要

[7] ISO 28004-1:2007, Security management systems for the supply chain − Guidelines for the implementation of ISO 28000 − Part 1: General principles

[8] CNS 29134 資訊技術－安全技術－隱私衝擊評鑑之指導綱要

[9] ISO/IEC 38500:2015, Information technology − Governance of IT for the organization

[10] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, De-Identification of Personal Information, NISTIR 8053. NIST, 2015

[11] ALLIANCE H.I.T.R.U.S.T. De-Identification Framework. HITRUST, 2015

[12] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, April 2014

[13] Branson J et al., Evaluating the Re-Identification Risk of a Clinical Study Report Anonymized Under EMA Policy 0070 and Health Canada Regulations' (2020) 21 Trials 200

[14] Data61 and the Office of the Australian Information Commissioner, De-identification Decision-making Framework, CSIRO Reports EP173122 and EP175702, 2017.

[15] European Medicines Agency, External Guidance on the Implementation of the European Medicines Agency Policy on the Publication of Clinical Data for Medicinal Products for Human Use' (2018) EMA/90915/2016 Version 1.4</unknown>

[16] INFORMATION COMMISSIONER'S OFFICE Anonymisation: Managing Data Protection Risk Code of Practice. Information Commissioner's Office, 2012

[17] Information and Privacy Commissioner of Ontario De-identification Guidelines for Structured Data, IPC of Ontario, 2016.

[18] APPENDIX B FROM INSTITUTE OF MEDICINE Sharing Clinical Trial Data: Maximizing Benefits. Minimizing Risk, 2015(1)

註(1) https://www.nap.edu/catalog/18998/sharing-clinical-trial-data-maximizing-benefits-minimizing-risk

[19] OFFICE OF NATIONAL STATISTICS, Guidance on Intruder Testing ((undated))(2)
註(2)
https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/guida nceonintrudertesting
[20] UK ANONYMISATION NETWORK, Anonymisation Decision-making Framework.UKAN Publications, 20163)
[21] Wagner I, Eckhoff D, Technical Privacy Metrics: A Systematic Survey' (2018) 51 ACM Computing Surveys (CSUR) 57:1

Bibliography

[1] ISO/IEC 23751, Information technology − Cloud computing and distributed platforms − Data sharing agreement (DSA) framework

[2] ISO/IEC/IEEE 24765:2017, Systems and software engineering − Vocabulary

[3] ISO/IEC 25012:2008, Software engineering − Software product Quality Requirements and Evaluation (SQuaRE) − Data quality model

[4] ISO/IEC 27002, Information technology − Security techniques − Code of practice for information security controls

[5] ISO/IEC 27050-1:2019, Information technology − Electronic discovery − Part 1: Overview and concepts

[6] ISO/IEC 27701, Security techniques − Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management − Requirements and guidelines

[7] ISO 28004-1:2007, Security management systems for the supply chain − Guidelines for the implementation of ISO 28000 − Part 1: General principles

[8] ISO/IEC 29134:2017, Information technology − Security techniques − Guidelines for privacy impact assessment

[9] ISO/IEC 38500:2015, Information technology − Governance of IT for the organization

[10] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, De-Identification of Personal Information, NISTIR 8053. NIST, 2015

[11] ALLIANCE H.I.T.R.U.S.T. De-Identification Framework. HITRUST, 2015

[12] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, April 2014

[13] Branson J et al., Evaluating the Re-Identification Risk of a Clinical Study Report Anonymized Under EMA Policy 0070 and Health Canada Regulations' (2020) 21 Trials 200

[14] Data61 and the Office of the Australian Information Commissioner, De-identification Decision-making Framework, CSIRO Reports EP173122 and EP175702, 2017.

[15] European Medicines Agency, External Guidance on the Implementation of the European Medicines Agency Policy on the Publication of Clinical Data for Medicinal Products for Human Use' (2018) EMA/90915/2016 Version 1.4</unknown>

[16] INFORMATION COMMISSIONER'S OFFICE Anonymisation: Managing Data Protection Risk Code of Practice. Information Commissioner's Office, 2012

[17] Information and Privacy Commissioner of Ontario De-identification Guidelines for Structured Data, IPC of Ontario, 2016.

[18] APPENDIX B FROM INSTITUTE OF MEDICINE Sharing Clinical Trial Data: Maximizing Benefits. Minimizing Risk, 2015[1)]

[19] OFFICE OF NATIONAL STATISTICS, Guidance on Intruder Testing ((undated))[2)]

[1)]https://www.nap.edu/catalog/18998/sharing-clinical-trial-data-maximizing-

benefits-minimizing-risk

[2] https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/guida nceonintrudertesting

[20] UK ANONYMISATION NETWORK, Anonymisation Decision-making Framework. UKAN Publications, 20163)

[21] Wagner I, Eckhoff D, Technical Privacy Metrics: A Systematic Survey' (2018) 51 ACM Computing Surveys (CSUR) 57:1

名詞對照

−A−	
adversarial testing	對抗性測試
adversary	對手
aggregated data	聚合資料
aggregation	聚合
agreement	協議
approach	作法
assessment	評鑑；評估
attribute	屬性
attribute disclosure	屬性揭露
attribute type	屬性型式
audit	稽核
audit trail	稽核存底
awareness	認知
−B−	
benchmark	效標
−C−	
compliance	遵循性
context	全景
control	控制；控制措施
controller	控制者
custodian	保管者
−D−	
data custodian	資料保管者
data de-identification process	資料去識別化過程
data principal	資料當事人
data privacy model	資料隱私模型
data recipient	資料接收者
dataset	資料集
data sharing agreement	資料分享協議
data transformation	資料轉換
data utility	資料效用
defined population	經定義母體

de-identification	去識別化
de-identification governance	去識別化治理
de-identification process	去識別化過程
de-identification technique	去識別化技術
de-identified data	經去識別化資料
direct identifier	直接識別符
disclosure	揭露
−E−	
event	事件
evidence	證據
−G−	
generalization	概化
governance	治理
granularity	精細度
guidance	指引
guideline	指導綱要
−I−	
identifiable uniques	可識別唯一項
identifiability	可識別性
identification	識別
identifying attribute	識別屬性
identity	身分；識別資訊
identity disclosure	身分揭露
incident	事件
information security	資訊安全
impact	衝擊；影響
implementation	實作；實施
indirect identifier	間接識別符
inference	推論
−J−	
journalist risk	新聞記者風險
−L−	
linking	鏈結

−M−

marketer risk	行銷人員風險
measure	度量
metric	指標
mitigation	緩解；緩解措施

−N−

noise addition	雜訊添加

−P−

penetration testing	滲透測試
personally identifiable information, PII	個人可識別資訊
PII controller	PII 控制者
PII principal	PII 當事人
PII processor	PII 處理者
policy	政策
population	母體
privacy	隱私
privacy by design	於設計即保護隱私
privacy impact assessment, PIA	隱私衝擊評鑑
privacy principle	隱私原則
probability function	機率函數
process	過程；處理
processor	處理者
prosecutor risk	檢察官風險
procedure	程序
profile	剖繪
pseudonym	假名
pseudonymization	假名化

−R−

record	紀錄
re-identification	重新識別
re-identification risk	重新識別風險
released data	經發布資料
responsibility	責任
review	審查
risk	風險

−S−	
sample	樣本
sensitive attribute	敏感屬性
shared data	經分享資料
singling out	單獨挑出
stakeholder	利害相關者
statistical twin	統計孿生
suppression	抑制
−T−	
threat	威脅
threat modelling	威脅建模
threshold	臨限值
tolerance	許可差
transparency	透明度
−U−	
uniqueness	唯一性
uniques	唯一項
utility	效用
−V−	
vulnerability	脆弱性
−W−	
written agreement	書面協議

相對應國際標準

ISO/IEC 27559:2022 Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework

中華民國國家標準
發行機關：經濟部標準檢驗局
局　　址：臺北市中正區濟南路一段四號
電　　話：(02)2343-1700
網　　址：https://www.bsmi.gov.tw
編輯排版：文山彩藝有限公司
銷售網址：https://www.cnsonline.com.tw
定　　價：依上開銷售網站公告之售價為準
GPN　　：4911300047