淺談個資「去識別化」與「合理利用」間的平衡

蔡柏毅 / 金融聯合徵信中心 法務室

一、前言

隨著科技的高速發展，關於個人資料的保護，以及資料主體資訊隱私權利的保障，一方面提高了資料遭受侵害的潛在風險，另方面也促進了資料的流通與利用。我國個人資料保護法（以下簡稱「個資法」）第 1 條立法目的即開宗明義揭示：「為規範個人資料之蒐集、處理及利用，以避免人格權受侵害，並促進個人資料之合理利用，特制定本法。」可知個資法立法目的一則在於保護個人資料，二則在於促進其合理利用。另歐盟個人資料保護規則（General Data Protection Regulation，以下簡稱'GDPR'）前言（recital）第4點亦表明：「個人資料之處理應為人類福祉而設，惟個人資料之保護並非絕對的權利，仍必須考慮其在社會上之作用，並與其他權利平衡兼顧。」

為衡平個資的利用及其風險，常見的立法例是透過某些方式或技術，移除或降低資料與資料主體之間的連結，使其無法識別或難以識別特定當事人，這個過程一般稱之為「去識別化」（de-identification）[1]，又可分為「排除個資適格性的去識別化」與「目的外利用的去識別化」，前者理論上已切斷資料與特定主體間的連結，後者仍有拼湊出個人圖像的更高可能性，因而對於資料主體潛在的威脅較高。

個資法第2條第1款規定個人資料之定義，除所列舉自然人之姓名等19項直接識別資料[2]外，亦例示包括「其他得以直接或間接方式識別該個人之資料」，依反面解釋，所謂去識別化即「無法以直接或間接方式識別特定當事人」之資料[3]，惟實際上要到什麼程度才算符

1 「去識別化」一詞並非法典用語，我國個資法及GDPR中雖然均以「識別」一詞作為個人資料在判斷上與操作上的核心概念，惟均未直接使用「去識別化」一詞，詳本文下述。

2 個資法第2 條第1款：「個人資料：指自然人之姓名、出生年月日、國民身分證統一編號、護照號碼、特徵、指紋、婚姻、家庭、教育、職業、病歷、醫療、基因、性生活、健康檢查、犯罪前科、聯絡方式、財務情況、社會活動及其他得以直接或間接方式識別該個人之資料。」

3 法務部103年11月17日法律字第10303513040號函釋：「去識別化之個人資料依其呈現方式已無從直接或間接識別該特定個人者，即非屬個人資料，自非個資法之適用範圍。」

合法規範所稱之「無法識別」，仍待進一步探究。

我國有關應用去識別化個資之代表案例，蓋為全民健康保險資料庫案，中央健康保險署將其所保有之國民納保與就醫之健保資料，經加密後提供予國家衛生研究院以建立健保研究資料庫，引發當事人重大權利爭議。該案終審判決被告（即今衛生福利部）勝訴[4]，惟法院於判決理由指出「去識別化應以完全切斷資料內容與特定主體間之連結線索」程度為準，該案資料處理者仍掌握還原資料與主體間連結之能力，其個資屬性並未全然排除，與去識別化之標準未符。最高行政法院同時強調去識別化之功能與作用，在於「確保社會大眾無法從資料內容輕易推知該資料所屬主體」，並有提到關於「資料可能被再識別」之風險評估，然而應採行何種評估標準，該次判決並未有明確說明。

二、GDPR 與去識別化相關之條文

摘錄GDPR前言及法典本文中與去識別化（含「匿名化」與「假名化」）相關之條文如下：

（一）前言（26）：

「本規則所定之各種個人資料保護原則應適用於可識別（identified）或可得識別（identifiable ）當事人之任何資訊。已假名化（pseudonymisation）之個人資料，其可透過運用額外資訊（additional information）而識別當事人身份者，應屬可得識別當事人的資訊。

為決定當事人是否可被識別，應考慮到所有可合理可能之方法（all the means reasonably likely to be used），例如可指認性（singling out）判斷，無論係由控管者自己或透過他人指認而直接或間接識別該特定當事人。為確認可合理用來作為識別當事人之方法，應考慮所有客觀因素，諸如：識別所需之成本、時間、資料處理當時之技術與科技發展等。

本規則所定之各種個人資料保護原則不適用於匿名（anonymous）資料，亦即非「已識別」或「可識別」當事人之資訊，或以使資料主體「不可」或「不再可」識別（not or no longer identifiable）之方法，而使其成為匿名之個人資料，包括為統計或研究目的所為之者。」

（二）前言（28）：

「對於個人資料應用假名化（pseudonymisation）技術，對於資料主體可降低風險，並可協助控管者及處理者履行其保護個人資料之義務。

惟本規則明確採用假名化一詞，並無意排除為資料保護目的所為之其他任何措施。」

4　最高行政法院106年判字第54號判決。有關本案之評介，可參閱所附「參考資料」中學者之相關著作，本文限於主題及篇幅無法詳述。

（三）前言（29）：

「為鼓勵於資料處理過程中應用假名化技術，控管者在被許可進行一般分析（general analysis）之情形，已採取必要之技術上及組織上措施（technical and organisational measures）以確保在處理過程中對本規則之遵循，且該得以識別特定資料主體之額外資訊已被分開存放（kept separately）者，假名化技術仍有其應用可能。」

（四）前言（57）：

「控管者處理之個人資料不允許其識別特定當事人時，不得單獨為遵循本規則任何規定之目的，為識別資料主體而獲取該額外資訊。但控管者不得拒絕接受資料主體為行使其權利所提供之額外資訊。

本規則所稱識別包括資料主體之數位辨識（digital identification），例如資料主體登入控管者所提供之網路服務時，所使用之憑證或認證機制等。」

（五）本文第4條「定義（Definitions）」：

(1)「個人資料」係指關於識別或可得識別自然人（「資料主體」‘data subject’）之任何資訊。可得識別自然人係指得直接或間接的識別該自然人，特別是參考諸如姓名、統一編號、位置資料、網路識別碼，或一個或多個該自然人之身體、生理、基因、心理、經濟、文化或社會認同等具體因素之識別符號（identifier）。

……

(5)「假名化」係指處理個人資料之方式，使該個人資料在不使用額外資訊時不再得以識別特定資料主體（no longer be attributed to a specific data subject without the use of additional information），且該額外資料已被分開存放，並以技術上及組織上的措施（technical and organisational measures）確保該個人資料無法或不再可識別該自然人。

（六）本文第6條「處理之適法性（Lawfulness of processing）」：

……

4.如處理係出於蒐集個人資料以外之目的，且非基於資料主體之同意，或非依據歐盟法或會員國法律在民主社會中為確保實現本規則第23條第1項所定目的之必要且適當方法（necessary and proportionate measure）所為時，為確保處理之目的與原先蒐集資料之目的兼容，應考慮包括但不限於下列事項：……

(e)適當保護措施之存在，包括加密（encryption）或假名化。

（七）本文第11條「不須識別之處理（Processing which does not require identification）」：

1.控管者處理個人資料之目的非為識別特定資料主體，或不再需要由控管者識別資料主體時，該控管者無義務維護、取得或處理以識別該資料主體為唯一目的之額外資訊。

2.如控管者得證明其非立於識別該資料主體之地位，該控管者應於可能範圍內通知該資料主體。於此情形，第15條至第20條規定[5]不予適用，但資料主體依該等規定，為行使其權利之目的提供得識別其身份之額外資訊者，不在此限。

（八）本文第32條「處理之安全性（Security of processing）」：

1.考量現有技術（the state of the art）、執行成本（the costs of implementation）、處理之本質（nature）、範圍（scope）、脈絡（context）及目的（purposes）與對當事人權利及自由的風險之可能性與嚴重性，控管者及處理者應採取適當之科技化且組織化措施以確保對於風險之適當安全程度，包括但不限於：

(a)個人資料之假名化及加密（the pseudonymisation and encryption of personal data）。……

三、「去識別化」、「匿名化」與「假名化」三個概念的釐清

「識別」在概念上分為「可識別」與「可得識別」二類。可透過與其他資訊（additional information）結合而識別當事人身份者，為間接識別；毋須透過與其他資訊結合即可識別當事人身份者，為直接識別。可識別即為「直接識別」，可得識別則屬「間接識別」。判斷一項資料或一個資料集（dataset）是否屬個人資料，一般係以「去識別化的程度」作為判準，換言之，去識別化是一個嘗試降低「被重新識別的風險」（re-identification risks）與保持資料使用於預定用途的可用性之間，找尋衡平的過程。已利用所有合理可能之資料去識別化方法，仍存在還原資料與資料主體間連結之可能性者，該資料仍為應受保護之個資；而如利用所有合理可能之方法，已「不可識別」或已「不再可識別」特定資料主體者，則屬完整去識別化之資料。

去識別化常伴隨被重新識別的剩餘風險（residual risk），因此僅能檢視是否已排除「合理可能」識別特定個人之程度，例如識別所需之成本是否過高、時間是否過鉅、當時的技術與科技的發展等。資料保有者須評估其所釋出之去識別化資料是否仍有合理可能使資料接收者得與其他資訊結合後，進而識別出特定個人。倘若釋出之資料本身雖不具特定個人識別性，但如以此為線索，與其他已公開、或一般人可得獲取之額外資訊組合、比對之後，亦得識別出特定個人時，該資料將仍屬得間接識別之個人資料。從而，所謂「其他資訊」、「額外資訊」之範圍及處理，亦將影響去識別化之判斷準據。

5 GDPR第15條至第20條為規範關於資料主體得以行使之權利，包括接近使用權（第15條）、請求更正及刪除權（被遺忘權）（第16條、第17條）、限制處理權（第18條）、受告知或通知權（第19條）及資料可攜權（第20條）等。

與去識別化相關，而更容易在定義與操作上被混淆的一組名詞是「匿名化」（anonymisation）、「匿名化資料」（anonymised data）與「假名化」（pseudonymisation）、「假名化資料」（pseudonymised data）[6]的區分，簡要說明其差異如下：

◎匿名化資料已非屬個人資料，假名化資料則仍受個資法保護：

匿名化資料須完全除去資料與特定主體之間的連結，已非屬個人資料；假名化資料則仍有可能透過連結其他資訊而間接識別當事之人身份，故仍受相關法律之規範及保護。

◎匿名化重視資料的「可識別性」，假名化則側重其「可識別程度」，進而決定個資「應受保護的規範密度」

匿名化過程係「移除資料的可識別性」，進而排除個資的適格性（變成不是個資）；假名化程度則為「資料可識別特定人的程度」，進而與資訊隱私權利應受保護的強度連動，亦即以去識別化為手段，以保障資訊隱私的規範設計，因此會進一步討論去識別化的程度，即個資應受保護的程度。

◎假名化資料可以匿名化，匿名化資料原則上不應退回到假名化：

假名化的過程會創建兩個資料集：假名化資料集及其「附加資訊」資料集。如附加資訊現實上已不存在，此時該假名化資料集因為無法再識別特定資料主體，事實上其性質已轉變為匿名化資料。然而，已匿名化之資料原則上不應重新變回假名化，蓋已匿名化資料如可隨時轉換回到假名化，則相當可疑，應可視為從未完成匿名化。

另一方面，匿名化資料亦可能隨時代演進而轉變其性質，舉例而言，過往曾廣泛被認為是匿名之定位資料（個人之移動軌跡）因具有高度相關性與獨特性，在某些情況下容易被重新識別出特定個人，需要進一步保護。[7]

◎假名化為匿名化工具之一，惟假名化的結果不一定是匿名化：

匿名化之程度與內涵應為固定判準，意即是否為個人資料必須清楚明確；但假名化作為匿名化（去識別化）的一種方式或手段，則存在程度不同之作業風險，唯有到達無從識別特定個人之地步，始為匿名化的完成。

◎匿名化的客體為資料整體，假名化則係就個別資料所為之處理：

個別資料在性質上無法被單獨匿名化，只有針對「資料整體」始有討論匿名化的可能，任何對個別資料片段之干預及處理（例如透過加密或其他轉換方式）均為假名化，而非匿名化。

6 「假名化」亦有譯為「擬匿名化」者，概取其相對於匿名化而非完全匿名化之意。

7 歐盟個資保護委員會（European Data Protection Board, EDPB）於2020年4月為面臨嚴峻疫情之公共衛生危機，發布「關於利用定位資料及追蹤接觸工具」之書面指引。

◎匿名化理論上為澈底的去識別化，假名化則是暫時去識別化：

匿名化資料在理想上須澈底去除識別特定個人的可能性；假名化資料則屬可間接識別之個人資料，結合其他資料即可還原為可直接識別之個資，因此性質上僅為暫時性的去識別化。

◎匿名化理想上已豁免風險，假名化僅是一種保護措施：

匿名化資料因已不再識別特定個人，理論上對資料主體已不存在風險；假名化的有效使用則能降低資料主體資訊隱私權利被侵害的風險，屬資料控管及處理者所應承擔的個資保護義務履行方式之一，亦為一種 GDPR有明文規定的「適當保護措施」（appropriate safeguards）。

例如因進行匿名化而留存「轉換程式」或「對照表」等額外資訊，應於事後將該轉換程式或對照表銷毀，或至少應將原始資料與匿名化資料分別由兩個不同機構保管，使被提供利用之資料與原始資料形成連結不可能之匿名化狀態，始為對資料主體之充份保護。

◎同一份資料可能同時是匿名化資料及假名化資料：

對資料的原始保有者而言，因通常持有得以將其「還原」或重新識別之額外資訊，故縱使對於資料的接收者或第三方（third parties）而言可能已經達到非屬個資的匿名化程度，惟對資料原始保有者而言仍屬間接可識別。因而視搜集者、處理者或利用者的角色不同，相同資料在同一時間可以既是匿名化資料，又是假名化資料。

我國個資法施行細則第3條規定之「間接識別」係指「須與其他資料對照、組合、連結等，始能識別特定個人」，與歐盟GDPR有關「間接識別」定義相近，仍屬個人資料。惟個資法施行細則第17條規定「無從識別特定當事人」係指「個人資料以代碼、匿名、隱藏部分資料或其他方式，無從辨識該特定個人」，此一規定如對照法務部103年函釋[8]所稱「去識別化之個人資料依其呈現方式，已無從直接或間接識別該特定個人者，即非屬個人資料」，似有混淆匿名化與假名化之誤解。

爰「代碼、匿名、隱藏部分資料或其他方式」仍可透過「與其他資料對照、組合、連結」以識別特定個人，仍屬間接識別之個人資料，因此個資法條文所稱「經提供者處理後或蒐集者依其揭露方式，無從識別特定當事人」之資料，性質上仍為一種間接識別性個資，亦即假名化個資，而不屬於個資法第2條定義「經處理後已無法再以直接或間接方式識別特定個人」之匿名化個資。蓋性質上已非屬個資

8 詳註3。

而不適用個資法之匿名化個資，其處理及利用本即無須再加上「基於公共利益為統計或學術研究而有必要」之客觀要件。

另「代碼、匿名、隱藏部分資料或其他方式，無從辨識該特定個人者」，其中「匿名」一詞係與「代碼」、「隱藏部分資料」併列，應為假名化方式之例示，而與本文所述之匿名化無關。

四、似是而非：匿名化並非資料利用的免死金牌

為決定是否可識別當事人，應考慮所有「合理可能」之方法（GDPR前言第26點），此處使用「合理」（reasonably）一詞，意指識別有其高低不同的成本與困難程度；而使用「可能」（likely）一詞解釋可識別性，代表仍存在「被識別」或「被重新識別」的概率與風險。

依資料與特定個人連結程度的不同，對資料主體造成資訊隱私風險的程度也隨之不同，就風險程度由高至低可分為「可直接識別」、「可間接識別」、「被識別風險較高而可能被識別」、「因被識別風險較低而不能被識別」及「無法識別」等五個階層（請參閱圖1）[9]。如前文所述，考慮所有合理可能之識別方法後，就無法識別當事人之有效匿名化資料，已排除個資適格而非屬個人資料，原則上毋需以法律加以規範及保護，惟如何界定資料已完成匿名化，隨著科技演進，其困難度與不確定性已日漸提升。

要實現絕對匿名（absolute anonymisation）是一項艱鉅的任務，匿名化資料必須是「不可逆」（irreversible；non-retraceable）的，匿名化資料客觀上應無法回復原始資料[10]。但在規範上，「不可逆」係指：透過「所有合理可能之方法」（而非以「任何方法」）均無法識別或再識別出當事人。由此可知，所謂「不可逆」的判準並非絕對，更無法真正做到僅存在理想上的「零風險」。

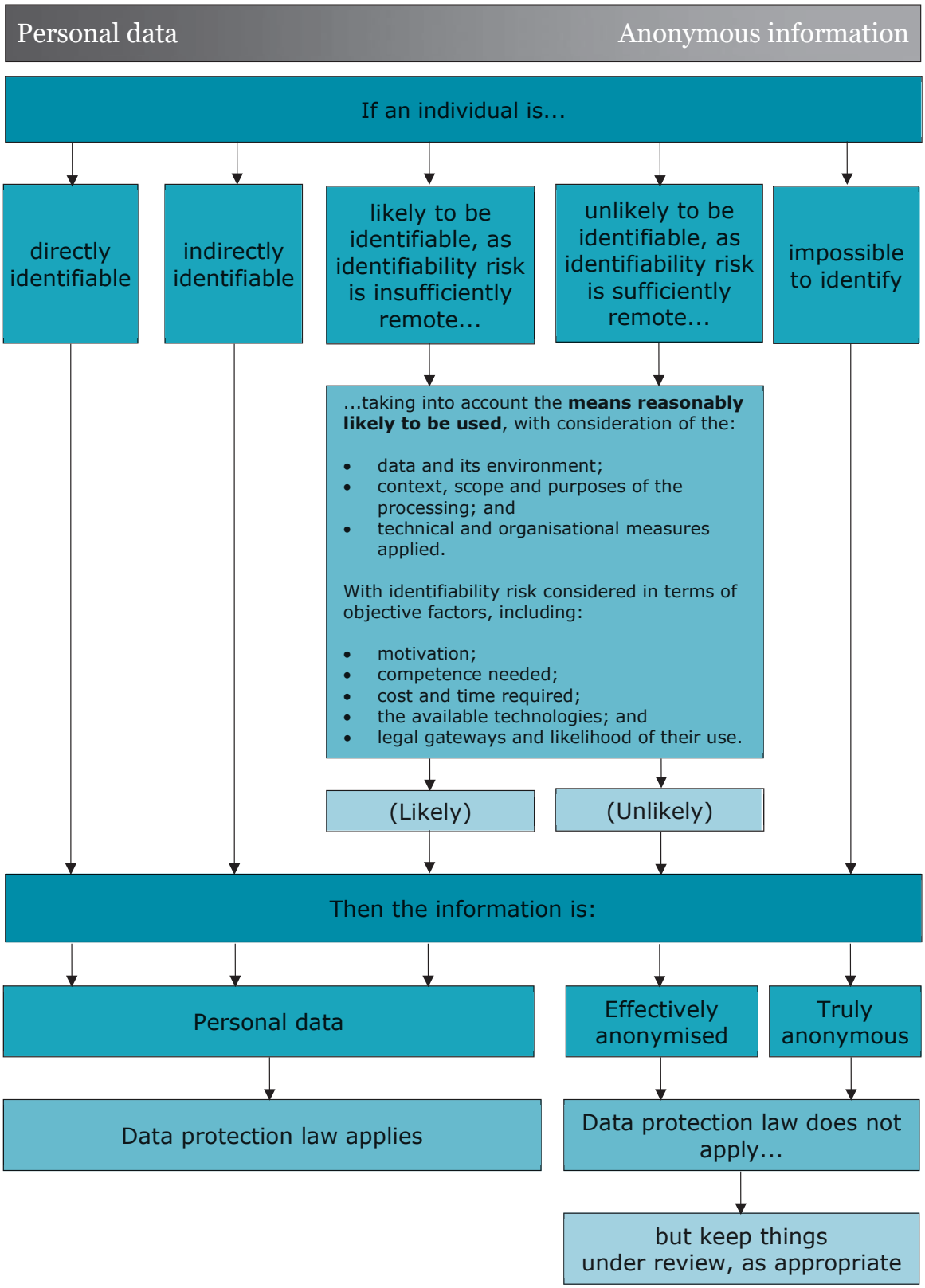
在技術層面上，諸如開放資料（open data）、資訊公開揭露（freedom of information）、資料探勘（data mining）、雲端運算（cloud computing）、乃至大數據/巨量資料的比較與參照（Big data comparisons）等技術層面的演進，各種資料大量的產生、然後被蒐集、連結與利用，其中必然包括具有極高市場價值的個人資料，甚至資料產業亦有被稱為「21世紀的黑金」者。[11]

9 圖1出處為英國個資主管機關即「資訊專員辦公室（Information Commissioner's Office, ICO）」於2021年10月8日公布之《匿名化、假名化及隱私強化工具指引》第二章之草案。前揭指引章節旨在釐清資料匿名化之概念，以及匿名處理的有效性判斷標準。

10 依法律規定允許重新識別之區域除外，例如：醫療實驗研究中，得以回溯追蹤受試病患，藉以調整對病患之醫療處置。

11 「21世紀的黑金」一詞原為法國社會經濟學家Guilhem Fabre對智慧財產權的形容，然而我們也可以說資料的處理與利用實為智慧財產權的重要基礎，姑假借於茲。

圖1


Personal data
Anonymous information
If an individual is...
directly identifiable
indirectly identifiable
likely to be identifiable, as identifiability risk is insufficiently remote...
unlikely to be identifiable, as identifiability risk is sufficiently remote...
impossible to identify
...taking into account the **means reasonably likely to be used**, with consideration of the:
• data and its environment;
• context, scope and purposes of the processing; and
• technical and organisational measures applied.
With identifiability risk considered in terms of objective factors, including:
• motivation;
• competence needed;
• cost and time required;
• the available technologies; and
• legal gateways and likelihood of their use.
(Likely)
(Unlikely)
Then the information is:
Personal data
Effectively anonymised
Truly anonymous
Data protection law applies
Data protection law does not apply...
but keep things under review, as appropriate

在前述資料的生命週期（life cycle）之中，因資料來源的廣泛性，以及龐大的資料數量，使其可供串連、比對的機會大幅增加，再加上自動辨識科技的發展、系統運算能力及儲存空間的大幅提升。縱使依當時科技與技術的發展，客觀上已足可被認為已完成匿名化的資料，隨著此類「再識別科技」（re-identification technology）的技術演進，仍可能透過與其他資料的比對、組合、連結而被重新識別，使資料越來越難以保持匿名化的安全狀態，因此嚴格來說，所有資料均應被認為是潛在的個人資料。

駭客界有此一說：未來的攻擊者，永遠比當下的攻擊者資源更多、技術更棒。更重要的是，攻擊的動機更強。小自一般性的鄉民肉搜（學術上稱為「對於特定資料主體所進行之不當資訊探究行為」），大到運用高端科技進行之「拼圖識別」（jigsaw identification），即令年代再久遠、再片斷而近似已無法復原之紀錄，均仍有可能藉助其他資訊來源（不論是否為公開）蒐集比對相關資訊，進而還原出特定當事人（事、物）的真容全貌。

不僅匿名化有前述問題，假名化技術亦然。以減縮資料識別特定主體至一定程度的假名化資料，仍可能遭遇暴力攻擊（brute force attack）而遭到破解，又或者透過結合各類別資料連結到同一個假名之下，或該假名使用的期間過長，因而使得以追溯到特定當事人的機率不正常的提高，均足使假名化掩蓋的效果近乎失效。

有鑑於對匿名化及假名化的誤解與濫用，西班牙個資主管機關即該國「資料保護局」（Agencia Española de Proteccion de Datos, AEPD）及歐盟資料保護監督機關（European Data Protection Supervisor, EDPS）於2021年4月聯名提出「十個對匿名化常見的誤解」書面指引，向公眾說明與釐清匿名化的概念，摘要如下：

誤解1：匿名化等於假名化

匿名化資料係指無法與特定個人相關連的資料，資料於完成匿名化之後無法識別個人，因此該資料可以不適用GDPR的規定。相對於匿名化的另一個概念是假名化，依照GDPR的定義，假名化資料係指該個人資料在不使用額外資訊時，不再能夠識別出特定之資料主體，且該額外資訊已被分開存放，並以技術上及組織上措施確保該個人資料無法或不再可識別出當事人。因此本質上，假名化資料仍屬於GDPR所保護的個人資料。

誤解2：加密等於匿名化

加密（encryption）不是一種匿名化的技術，而是屬於假名化的工具。加密過程中透過資料轉換，降低資料遭不當存取的風險，並保持一定的機密性。加密可以用解密金鑰（decryption key）加以還原，原始資料於解密之後可以被存取，此時解密金鑰即屬於前述「額外資訊」，進而識別特定個人。考量加密演算法、金鑰的強度與技術的進步（如：量子運算quantum computing）等因素，縱使將金鑰刪除仍無法確保資料無法再被解密，故無法認為加密資料即為匿名化資料。

誤解3：資料總是可以匿名化

匿名化是一個在「降低被重新識別的風險」及「保持資料的可用性」間找尋正確平衡的過程。然而根據資料的性質及使用情境，有時候無法充分降低被重新識別的風險，例如資料的總數過小時（如一個僅包含705名歐洲議會成員的匿名資料集），或是各個資料主體間的資料差異性特別時（例如GPS位置資料），均有可能無法充分降低被重新識別的風險。

誤解4：匿名化是永久的

實際上，隨著時間演進，不論是因為科技日新月異的進步（雲端運算與量子運算等技術），或儲存多年的資料發生外洩，均可能將過去已完成匿名化的資料重新識別到個人，進而喪失其匿名化屬性。

誤解5：匿名化資料被重新識別的風險趨近於零

匿名化的過程、使用的技術及應用的方式，對重新識別風險發生的可能性有直接的影響。強而有力的匿名化目的在將被重新識別的風險降低到特定閾值以下，而該閾值取決於現有的控制措施、被重新識別對資訊隱私的影響、攻擊者的動機及重新識別的能力等等，因此現實上被重新識別的剩餘風險並不總是為零，除非在資料的性質確實非常概化的特定情形下（如一年中每個國家或地區的網站存取者資料統計數據）。

誤解6：匿名化是絕對的概念且無法被衡量

匿名化資料不能被單純理解為匿名與非匿名的二維概念（binary concept），相對的，對匿名化程度進行分析與衡量是可能的。任何資料都必須依其個人化的可能性分別進行判斷，並評估其被重新識別的風險，同時必須進行常態性的管理與控制措施。

誤解7：匿名化可以完全自動化

匿名化的過程中須要對原始資料集、預期的目的、應用的技術與被重新識別的風險進行分析，除去或遮蔽直接識別資料是匿名化過程中的重要部分，但必須同時尋找並謹慎分析其他可能的識別來源。因此儘管匿名化的過程可能透過自動化工具完成，但專家的人工介入仍有其必要性。

誤解8：匿名化使資料喪失可用性

匿名化的目的是防止個人被識別出來，例如將自然人的出生日期以年份進行分組，因此匿名化在某些情形下確實降低了資料的可用性，但這並不意味匿名化資料毫無用處，其可用性取決於重新識別的目的與其風險。GDPR「資料最小化原則」要求資料控管者於進行資料處理前確認其必要性，因此在某些情形下確實可能得出匿名化資料不符合預期目的之結論，此時控管者必須在處理個人資料與遵循GDPR之間作出選擇。

誤解9：遵循其他成功的匿名化過程，就可以成功將資料匿名化

匿名化與做菜並不相同，就算完全按照食譜步驟操作仍然可能做出風味各異的菜色，更何況匿名化必須考量包括資料的性質、範圍、背景與目的、資料主體權利與自由之不同可能

性及不同程度的風險等複雜的變因。例如在瑞典，納稅人的個人資料是公開資料，在西班牙則否，因此一個包含西班牙與瑞典公民資料的資料集，縱使依照相同程序進行匿名化處理，對各別風險的判斷也可能不同。

誤解10：識別資料歸屬於何人，不存在風險或利益的考量

資料本身即具有價值，重新識別資料主體的身份將可能對個人的權利與自由發生嚴重的影響，而該影響始終取決於情境與相關資訊的判斷。例如透過對電影的偏好進行識別，可能據以推斷個人的政治傾向或性傾向，然而政治傾向或性傾向在GDPR中屬於敏感性個人資料，因而受到特別保護。

五、匿名化技術與資料的利用

依歐盟個人資料保護指令（Data Protection Directive 95/46/EC）第29條設立之「個資保護工作小組」（Article 29 Data Protection Working Party, WP29）[12]於2014年4月發布之書面指引，就資料的利用與再利用（secondary use），判斷匿名化技術有效性的標準包括：

◎指認性（singling-out）：

指透過觀察與分析，將特定個人自資料母群體(data principal)中挑出、分離並確定其特徵。

◎連結性（linkability）：

指連結屬於同一個人或同一組人的兩個以上資料集的紀錄，或連結不同資料集中的同一組資料主體。

◎推論性（inference）：

指使用其他資訊進行推斷、猜測或估計，有相當概率可以從某一組屬性(attribute)準確推斷出另一組屬性。

匿名化技術需具備上述標準特徵：可以有效避免藉由連結相關資料集，分離並挑出特定個體的紀錄，並透過推論方法加以辯認。茲以學界認為最適合結構化資料格式、易於執行風險評估、且經過驗證能在重新識別風險與保留資料利用性（Utility）間取得平衡之「K-匿名理論（K-anonymity）」為例，美國學者Latanya Sweeney的知名研究表明，僅僅使用三項間接識別資料（indirect identifiers），即「生日」、「性別」及「五位數郵遞區號」（ZIP/postal Code），即可由資料集準確識別87% 的美國人口。

上述Sweeney的研究背景，實基於一件重大的去識別化漏洞所導致的隱私漏洞案件：美國麻州於1996年將州政府員工的醫療資料於去識別化之後，提供公共醫學研究使用。當時就讀於麻省理工學院（MIT）的博士生

12 歐盟個資保護指令第29條工作小組（WP29）於GDPR施行後改制為「歐洲個人資料保護委員會（European Data Protection Board, EDPB）」，職責包括發布關於個資保護指令及GDPR的指引文件，並進一步闡釋或舉例說明相關條文之具體適用情形。

Latanya Sweeney使用連結攻擊法（linking attack），僅使用患者的三項屬性紀錄即破解前述資料集，順利找出時任麻州州長的醫療紀錄，並將之寄給州長本人。此一發現後來並影響美國「健康保險可攜與課責法」（Health Insurance Portability and Accountability Act，HIPAA）中「安全港（Safe Harbor）隱私規則」的建立。

假名化資料相較於匿名化資料，亦面臨類似的重新識別風險。首先，數據的洩露可能使得攻擊者獲得解密金鑰進而還原資料，或可能以其他手法將假名資料與個人身份連結起來（linkability）。或者即使金鑰沒有洩露，惡意行為者也可以通過將假名資料集中的間接標識資料與其他可用資訊結合，藉由推論以識別個人（inference）。

為在確保資料主體權利的前提之下，進行資料的共享與利用，以平衡風險並保留資訊的可利用性，有可能藉由「受信任的第三方」（trusted third parties, TTP）作為資料原始持有者與尋求使用資料者之間的中介，即由可信的第三方就原始資料實施匿名化或假名化技術後，再將資料提供予需求方，並以具公信力的認證機制及締結有效拘束力的契約等方式，使原始資料處於持有者的有效控管之下。

另一個可用的方式是運用「合成數據」（synthetic data）或稱「加鹽」（salting），是指在資料集的樣本數據中添加其他新數據，但不刪除或重新創造任何直接識別資料[13]。一方面可以保留該資料集之中各項資料的相關性，另方面可以在避免被直接識別的情況下從資料集中發現所需的趨勢與分析結果。

最後，就匿名化技術的選擇方面，亦可參用相關國家標準或國際標準，例如國際標準組織（International Organization for Standardization, ISO）發布之ISO 25237:2017「使用假名化服務以保護個人健康資訊的隱私保護原則與要求」，又或經濟部標準局CNS 29191「資訊技術－安全技術－提供個人資料部分匿名及部分去連結鑑別之要求事項」等。

六、代結論

本文所討論的「無法識別」、「無從被再識別」概念，最後終究要回到個資法所欲保護的「資料主體」自身去討論方有實益，茲舉一例，以代結論：

「某甲洗澡時，忽然有陌生人闖進來，某甲手上只有一塊小方巾，不知道要遮那邊，想一想，不如就遮臉好了。理由是：至少被看光光之後，人家不知道是誰光著身子。」

13 直接識別資料（direct identifiers）係指可直接識別個人的資料，可參閱註2我國個資法第2 條第1款規定列舉之19 項資料。另參照美國健康保險可攜與課責法（HIPAA）之定義，直接識別資料共有18 項，包含姓名、電話號碼、email 帳號及其他特徵或代碼等。

政大法律系教授劉宏恩在課堂上用此例問學生：有人願意遮臉裸身嗎？沒有學生願意。他認為，這就是現在常見的見解：拿掉姓名、身分的識別，就以為是「去識別化」。

但問題是：「把臉遮住，你就願意被看光光了嗎？」

參考文獻：

一、專書

1.Maria Cristina Caldarola、Joachim Schrey著，趙彥清、黃俊凱譯，大數據與法律實務指南（原著：Big Data und Recht: Einführung für die Praxis），元照，初版，2020年6月。

2.張陳弘、莊植寧，新時代之個人資料保護法制：歐盟GDPR與臺灣個人資料保護法的比較說明，新學林，初版，2019年6月。

二、期刊論文

1.L. Sweeney, Simple Demographics Often Identify People Uniquely, Carnegie Mellon University, Data Privacy Working Paper 3., Pittsburgh 2000.

2.江耀國，個人資料的概念與匿名化：一個認識論的觀點，東海大學法學研究，第58期，2019年9月。

3.李寧修，個人資料合理利用模式之探析：以健康資料之學術研究為例，國立臺灣大學法學論叢，49卷1期，2020年3月。

4.邱文聰，被淘空的法律保留與變質的資訊隱私憲法保障—評最高行政法院一〇六年度判字第五四號判決與相關個資法條文，月旦法學雜誌，272期，2018年1月。

5.吳全峰、許慧瑩，健保資料目的外利用之法律爭議—從去識別化作業工具談起，月旦法學雜誌，272期，2018年1月。

6.范姜真媺，個人資料保護法關於「個人資料」保護範圍之檢討，東海大學法學研究，第41期，2013年12月。

7.林其樺，個人資料保護、利用與匿名化—歐盟法制趨勢觀察，科技法律透析，28卷6期，2016年6月。

8.林其樺，數位浪潮：由歐盟個人資料管理制度與英國匿名化探索資料合理使用，科技法律透析，29卷1期，2017年1月。

9.林裕嘉，我國個人資料去識別化法制及實務發展概述，科技法律透析，28卷6期，2016年6月。

10.林裕嘉，公務機關利用去識別化資料之風險評估及法律責任，司法周刊，1852期、1853期，2017年6月。

11.郭戎晉，日本個人資料保護法修正重點與去識別化推動剖析，科技法律透析，28卷6期，2016年6月。

12.蔡昀臻、樊國楨，大數據之資料去識別的標準化實作初探：根基於 ISO/IEC 2nd WD 20889，資訊安全通訊，22卷4期，2016年5月。

三、歐盟資料

1.Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, 01248/07/EN WP136.

2.Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN WP216.

3.AEPD、EDPS, 10 Misunderstandings related to Anonymisation, last retrieved on December 15. 2021, from: https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en

4.EDPB, Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak, adopted on 21 April 2020.

5.ICO, Chap.2 How do we ensure anonymisation is effective?, Anonymisation, Pseudonymisation and Privacy Enhancing Technologies Guidance (draft), last retrieved on December 15. 2021, from: https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf

四、網頁資料

1.梁玉芳、彭慧明，去識別化只是用小毛巾遮臉，聯合報願景工程專題「隱私網戰」，2019年3月21日，https://vision.udn.com/vision/story/12932/3720706。

（最後瀏覽日期：2021年12月15日）