論 著
© 臺大法學論叢 NTU Law Journal
第 52 卷第 3 期/Vol.52, No.3 (09. 2023)

個人資料之去識別化與再識別化風險：法律之觀點*

翁清坤**

<摘要>

當前盛行以大數據技術進行資料研究分析，導出各種創新性推論或發現，以造福社會。惟資料廣泛運用時，亦漸形成隱私風險。

面對隱私保護與資料效用的衝突，倘將個資去識別化、匿名化，則不受個人資料保護法拘束，可移作原始蒐集目的外之利用或與第三人分享，以供各種運用。

惟大數據時代，有眾多資料來源可供交叉比對，不論去識別化或匿名化資料均難以維持不可逆、不可還原的狀態，而不可避免均有被再識別化風險，乃形成隱私等人格與經濟損害、表意自由的寒蟬效應。

一些知名的再識別化事件，使得去識別化的有效性漸受質疑。但有反駁，被再識別出來的比例實屬微小，去識別化仍屬有效機制。對此，美國法院見解亦相當分歧。

對於去識別化與再識別化的衝突，建議可採下列因應措施：（一）去識別化資料仍可能與其他資料相結合而再識別化，故較務實解決之道，應非在於完全排除再識別化風險，而應著重於減緩風險至極低程度。類似此風險忍

* 作者誠摯感謝匿名審查人之寶貴意見與費心指正，惟文責皆由作者自負。本文係國科會專題研究計畫之執行成果（MOST109-2410-H030-055 ）。

** 輔仁大學財經法律系專任副教授，美國威斯康辛大學麥迪遜分校法學博士。
E-mail: 071617@mail.fju.edu.tw

・投稿日：05/07/2022；接受刊登日：03/29/2023。
・責任校對：高映容、黃品樺、辛珮群。
・DOI:10.6199/NTULJ.202309_52(3).0001

受概念，歐美許多立法例普遍採用之「合理」識別化、去識別化標準，亦未要求「完全排除被再識別化之風險」。（二）去識別化的進行，應按再識別化風險評估而兼採符合比例之合理技術、行政與法律措施，以降低再識別化風險。（三）課予民刑事責任而禁止不當再識別化。

關鍵詞：個人資料、隱私、大數據、識別符號、去識別化、匿名化、再識別化、表意自由、風險忍受、一般資料保護規則

♦目 次♦

壹、前 言

當前大數據與人工智慧盛行，數位經濟與社會發展常奠基於規模前所未見的龐大個人資料（下稱「個資」）蒐集之上，包括個人搜尋、瀏覽紀錄、社會關係、醫療史[1]等在內涉及個人身分、行為、喜好的資料均被廣泛蒐集，並可與各產業、政府機關或研究人員分享。個資的蒐集、處理、利用與分享固可滿足各種需求而造福大眾，但亦將引起隱私侵害疑慮。

「個人資料」概念，歷來做為資訊隱私規範基石，在各國個資保護規範架構下似乎可行。不論何種資料，只要牽涉一個已被識別或可資識別出之自然人（any information relating to an identified or identifiable individual）[2]，即啟動隱私保護機制。

以大數據進行分析的這一年代，無疑將成就許多新興事物。大數據運作，藉由新穎的衍生性第二次利用（secondary use）、再利用（reuse）[3]，從原始資料集（dataset）萃取、推衍出隱藏價值。惟個資再利用常超出原始蒐集目的之利用，將與個人資料保護法（下稱「個資法」）之「目的限制原則」（principle of purpose limitation）有所衝突。在保護個人隱私與善用大數據之利益間，如何取捨平衡，乃一大挑戰。解決之道之一[4]，即將個資轉化為去識別化（de-identification）、匿名化（anonymization）[5]資料而不再符合個資

---

[1] Arvind Narayanan & Vitaly Shmatikov, *Privacy and Security: Myths and Fallacies of "Personally Identifiable Information"*, 53 COMM. OF THE ACM 24, 24 (2010). https://www.cs.utexas.edu/~shmat/shmat_cacm10.pdf (last visited Apr. 14, 2022).

[2] Omer Tene, *Privacy Law's Midlife Crisis: A Critical Assessment of the Second Wave of Global Privacy Laws*, 74 OHIO ST. L.J. 1217, 1239 (2013).

[3] 再利用與第二次利用在本文指同一概念而交替使用，未刻意加以統一，以尊重文中各處所引用不同文獻之原始用語。

[4] Sophie Stalla-Bourdillon & Alison Knight, *Anonymous Data v. Personal Data-A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data*, 34 WIS. INT'L L.J. 284, 285 (2016).

[5] 一般認為，個資「去識別化」與「匿名化」之定義，均著重於強調連結性之去除、或直接、間接個人識別符號之去除；有主張，匿名化乃去識別化之子類別

定義，從而不在隱私法拘束範圍[6]，乃可移作原始蒐集目的外進一步利用或分享予原本無權近用之第三人。在完全不與過多釋出個資之間，「去識別化」被視為一完美的妥協、折衷因應方法[7]。去識別化之目的，在於允許利用資

（subcategory）。惟二者差異之處，乃去識別化之方式未特別要求必須不可逆、不可還原的，因此，去識別化資料有可能因「利用某一鑰匙（如代碼、假名、演算法）而再連結回去至個人」；惟對於匿名化之方式是否不可還原的而絕對不能有再識別化可能性（即再識別化風險為零），則見解分歧。其實，對於再識別化的可能性、機率，聚訟盈庭。對於匿名化之方式是否限於不可還原，我國憲法法庭111年憲判字第13號判決採不可還原的見解，但仍有不同意見（參黃昭元大法官部分不同意見書第20段、本文註418）。

「去識別化」與「匿名化」二者名稱縱不相同，但定義方式卻有往同方向發展的趨勢，而均有要求以「合理可能」之識別手段為限。一方面，既然去識別化資料不免仍殘留再識別化之風險，乃有立法例（如美國法）要求應有「去識別化資料不再具有識別某一個人之合理可能性」之要件；另一方面，歐盟資料保護指令（DPD）第29條工作小組（Article 29 Working Party, WP29）指出，匿名的資料乃指「在考量所有合理可能用以識別該個人之手段而仍未能被識別出來」之任何涉及某一自然人之資料，詳述如本文「貳之三」單元。綜上，二者有視為不同概念，但亦有學者、組織，甚至主管機關視為同義詞而交替使用。LUK ARBUCKLE & KHALED EL EMAM, BUILDING AN ANONYMIZATION PIPELINE 6, 9 (2020); SIMSON L. GARFINKEL, DE-IDENTIFICATION OF PERSONAL INFORMATION 2 (2015), http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf (last visited Nov.30, 2022). 再者，由於本文根本立論基礎，乃「大數據時代，有眾多資料可供交叉比對，不論去識別化或匿名化資料，不論定義上是否要求去連結後需不可還原的，卻殊途同歸而不可避免均有被再識別化之風險」；加上本文所引用許多英文文獻常未區分二者之不同，而未將「匿名化」限定於絕對不可還原之情形，反而常將「匿名化」意涵視為等同於「去識別化」。因此，除在本文「貳之三」單元關於定義探討而介紹某些「二者主要因是否不可還原之差異」而係不同概念之論述外，原則上，本文行文通常亦將「去識別化」與「匿名化」視為同義詞而交替互用，但其在各處之具體使用常亦兼顧所引用文獻原始用語。

[6] Mike Hintze, *Science and Privacy: Data Protection Laws and Their Impact on Research*, 14 WASH. J. L. TECH. & ARTS 103, 114 (2019);美國加州隱私權法（California Privacy Rights Act of 2020）、維吉尼亞州消費者資料保護法（Consumer Data Protection Act）、科羅拉多州隱私法（Colorado Privacy Act）及統一法律委員會（Uniform Law Commission）所起草的統一個人資料保護法（Uniform Personal Data Protection Act，下稱UPDPA）均明定，個資不含去識別化資料。

[7] Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J. L. & TECH. 1, 67 (2011); Kelsey Finch & Omer Tene, *Welcome to the Metropticon: Protecting Privacy in A*

料時，仍可藉由屏蔽資料主體（當事人）資料、身分而提供某種隱私保護避風港。基於隱私考量而本來被禁止利用之資料，因去識別化而可迴避風險與法遵要求，進而開啟嶄新、第二次利用[8]的可能，以增進資料運用所能帶來的進步與福祉。

惟近來研究顯示，大數據環境中，縱去除個資，已被去識別化、匿名化之資料仍可能與散落他處之資料來源互相結合比對，而難以絕對確定「一個人不可能由某一特定資料集而被識別出來」[9]；由於存在再識別化（re-identify）、去匿名化（deanonymization）的風險，去識別化、匿名化資料不再能不可逆（轉）、不可還原地（irrevocably）免於重新被識別出來。因此，去識別化、匿名化概念飽受質疑，對其做為隱私保護之有效性、可能性，法律、隱私學者與專家見解相當分歧[10]，法院判決亦分歧。

其中，一派學者質疑，個人資料或個人可識別資料（personally identifiable information，PII）概念儼然已成隱私的一種魔法石（lapis philosophorum）、萬靈丹；因許多實務人士信奉「藉由去除（removing）或修改（modify）PII，即可將含有敏感個資的檔案加以去識別化」，如同中世紀煉金術士深信能將

Hyperconnected Town, 41 FORDHAM URB. L.J. 1581, 1613 (2014).

[8] THE DHHS OFFICE FOR CIVIL RIGHTS (OCR), GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA) PRIVACY RULE 5, https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard (last visited Apr. 14, 2022); Jules Polonetsky, Omer Tene & Kelsey Finch, *Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification*, 56 SANTA CLARA L. REV. 593, 594 (2016).

[9] President's Council of Advisors on Science and Technology (PCAST) (May 2014), *Big Data and Privacy: A Technological Perspective* 38-39, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf (last visited Apr. 14, 2023); C. Christine Porter, *De-identified Data and Third Party Data Mining: The Risk of Re-identification of Personal Information*, 5 SHIDLER J. L. COM. & TECH. 3 (2008).

[10] Elizabeth A. Brasher, *Addressing the Failure of Anonymization: Guidance from the European Union's General Data Protection Regulation*, 2018 COLUM. BUS. L. REV. 209, 231-32 (2018).

魔法石變成金子[11]。由於去識別化（匿名化）資料仍保留某些得藉以識別之風險，縱風險很小，仍非等於零，仍有可能連結回去資料本所對應之當事人；隱私保護與資料效用不相容，不易兼容並蓄。另一派學者則持肯定見解而主張，去識別化（匿名化）雖具侷限性，縱經極端努力而被再識別化風險仍非常低；某些專家誇大了再識別化風險之危害性而低估資料對外釋出的價值；其實，去識別化仍能有效減緩資料效用與隱私風險間的緊張關係。

對於去識別化資料之被再識別化風險，以及對於去識別化有效性的分歧見解，究竟可採取因應措施為何?深值探究，以求藉由去識別化技術釋出有用資料而促進公私領域的研究分析，仍能減緩隱私侵害。

因此，本文將探討與釐清歐美國家關於「個人資料之去識別化與再識別化風險」議題之相關法律論述，以供我國資料保護與運用[12]相關規範省思及設計之借鏡。本文在結構上，第貳部分探討個資之去識別化，第參部分分析去識別化資料之再識別化風險，第肆部分分析學界與實務界對於去識別化、匿名化有效性之分歧見解，第伍部分探討去識別化與再識別化衝突的可能因應之道，第陸部分則為結論。

貳、個人資料之去識別化

去識別化，乃用以去除可識別個資的一種工具；當政府、產業欲將資料供其他目的或外部人取用而開啟各種嶄新、第二次利用，去識別化顯得格外重要。關於去識別化之重要功能、應否取得同意、定義與方法，分述如下：

---

[11] Narayanan & Shmatikov, *supra* note 1, at 10.

[12] 如「臺北市大數據中心」的成立，以求有效整合不同數據資產並輔助施政決策，邁向智慧城市。臺北大數據中心網站，https://tuic.gov.taipei/（最後瀏覽日：11/30/2022）。

一、去識別化之重要功能

（一）去識別化資料不再符合個資定義而免受拘束而可供多元運用

當前各產業與政府部門正日益蓬勃地蒐集、處理、利用巨量資料（包括個資與非個資在內），藉由大數據進行各種創新性研究，以導出某些推論或發現，滿足社會各種需求。尤其，資訊公開、開放資料（open data）大纛之下，近來政府與大型組織漸公開釋出大量資料[13]，以促進公益、學術研究及產業發展[14]。如健康資訊不僅能改善醫療品質與效率予個人，尚有益醫學研究、公衛等。

任何含有可識別個人的資料蒐集、處理、利用或分享，皆應遵守個資保護原則（如取得同意），以防隱私侵害。惟隱私規範不免妨礙資料流通自由，不利政經社會運作[15]。當資料具個人可識別性而應受隱私規範，欠缺可識別性則可擺脫法律義務[16]。當原始資料本含有個資，經去識別化後[17]，不再符

---

[13] 例如，美國學者武雅士（Arthur P. Wolf）自1960年代起，分析臺灣日治時期海山郡3萬多筆戶口資料，發現童養媳婚的生育率普遍偏低、離婚率較高、妾婚率也較高；應證了芬蘭人類學家Edward Westermarck的「性嫌惡理論」（sexual aversion）：若兩性自小共同生活，長大後通常不會對彼此來電。其後，中央研究院所建置「日治時期戶口調查」資料庫（https://www.rchss.sinica.edu.tw/popu/index.php［最後瀏覽日：11/30/2022］），遵循去識別化原則，地址以代碼呈現並去除番地（街區號碼）訊息，保護個資，供學者申請使用。參自由時報（11/30/2022），〈解開臺灣歷史上的人口謎團！專訪中研院歷史人口研究計畫〉，https://talk.ltn.com.tw/article/breakingnews/4140215（最後瀏覽日：11/30/2022）。

[14] IRA RUBINSTEIN, BRUSSELS PRIVACY SYMPOSIUM ON IDENTIFIABILITY: POLICY AND PRACTICAL SOLUTIONS FOR ANONYMISATION AND PSEUDONYMISATION: FRAMING THE DISCUSSION 2, https://fpf.org/wp-content/uploads/2016/11/Rubinstein_framing-paper.pdf (last visited Apr. 14, 2022).

[15] Brasher, *supra* note 10, at 217.

[16] Tene, *supra* note 2, at 1239.

[17] GARFINKEL, *supra* note 5, at 1.

合個資定義[18]，則不在隱私法拘束範圍[19]。尤其，不再受個資規範之目的限制原則拘束，較易以新穎、多元方式而利用去識別化資料[20]，尚可安全分享資料供各種目的運用[21]。

去識別化（匿名化）資料是前個資（ex-personal data），不再能由資料重新識別出個人。非個資（如天候或大眾運輸時刻）則不含個資。如同完全公開的資料，非個資常可不受任何限制而釋出，惟未經處理的原始個資則不應完全開放而釋出。理論上，去識別化資料可如同公開的資料沒有限制再利用之必要而加以揭露[22]。去識別化資料堪稱社會發展基石，可以各面向造福大眾，從教育計畫、交通與都市規劃、反貪腐、基因研究、醫療[23]、到科技防疫[24]皆屬之。

---

[18] 其實，「個資」定義與「匿名化」定義二者堪稱一體正反二面。其中，針對「個資」定義，歐盟乃採合理審查標準（reasonableness test）。按歐盟GDPR Recital 26規定，「為判斷某一自然人是否可識別的，必須考量資料控管者或任何其他人所有合理可能（reasonably likely）用以直接或間接識別（如挑選出）某自然人之所有手段。」此「合理可能用以識別之手段」的審查標準不僅適用於「個資」判斷，亦用於評估「匿名化過程是否足夠完備」，即識別是否已成為合理地不可能。倘須投入不成比例的時間、費用與人力，才能識別，則不符上開個資定義；因此，縱因可能借助不成比例手段而仍存有被再識別化風險，但只要其低於可接受的合理風險門檻、檻值，即屬有效的匿名化（去識別化）而不再屬於個資，則不受隱私法拘束，故不再以取得同意為必要，而即可供原始蒐集目的外之利用。詳見本文貳之三之（二）之2之（1）之A、伍之一之（一）之3、伍之一之（二）之1單元。

[19] Hintze, *supra* note 6, at 114.

[20] THE INFORMATION COMMISSIONER'S OFFICE (ICO), ANONYMISATION: MANAGING DATA PROTECTION RISK, CODE OF PRACTICE 8 (2012), https://ico.org.uk/media/1061/anonymisation-code.pdf (last visited Apr. 14, 2022).

[21] RUBINSTEIN, *supra* note 14, at 2.

[22] Frederik Zuiderveen Borgesius, Jonathan Gray & Mireille van Eechoud, *Open Data, Privacy, and Fair Information Principles*: *Towards a Balancing Framework*, 30 BERKELEY TECH. L.J. 2073, 2130 (2015).

[23] RUBINSTEIN, *supra* note 14, at 2.

[24] 例如，根據臺灣2,300萬人、6千多萬劑COVID-19疫苗之去識別化後資料的科學分析，打疫苗者比起未打疫苗者，可減少中重症及死亡發生；無論是哪種組合（高端、莫德納或BNT疫苗混打），接種3劑疫苗者，避免中重症及死亡保護效益都優於未打滿3劑疫苗者。自由時報（11/18/2022），〈臺灣疫苗混打真實世界數據出

財務考量，亦常是資料去識別化的另一動機。例如，個資外洩通知成本過高，預估每一受影響個人為 200 美金，大型資料庫更將累計可觀金額；惟資料去識別化後，毋庸進行外洩通知，可避開高昂成本[25]。甚至，某些業者（如資料仲介公司）在去除可識別個人的資料後，售予他人[26]，供更多元運用。

（二）各國立法納入去識別化概念而在隱私保護與資料效用間尋求平衡

去識別化（匿名化）概念深植規範政策，某些隱私法規即納入去識別化概念而視其為一種可符合法律要求的方法[27]，可放寬管制而移作他用，以在隱私保護與資料效用間達成最佳平衡[28]。國際上相關立法例，如我國個資法（如第 6 條第 1 項第 4 款[29]、第 16 條第 5 款[30]、第 20 條第 1 項第 5 款[31]）規

爐　高端混打 BNT 防中重症效力最高〉，https://health.ltn.com.tw/article/breakingnews/4127698（最後瀏覽日：11/30/2022）。

[25] KHALED EL EMAM & LUK ARBUCKLE, ANONYMIZING HEALTH DATA 3 (2013).

[26] Theodore Rostow, *What Happens When an Acquaintance Buys Your Data?: A New Privacy Harm in the Age of Data Brokers*, 34 YALE J. ON REG. 667, 688-89 (2017).

[27] GARFINKEL, *supra* note 5, at 4-5.

[28] RUBINSTEIN, *supra* note 14, at 2.

[29] 個資法第6條第1項第4款：「有關病歷、醫療、基因、性生活、健康檢查及犯罪前科之個人資料，不得蒐集、處理或利用。但有下列情形之一者，不在此限：……四、公務機關或學術研究機構基於醫療、衛生或犯罪預防之目的，為統計或學術研究而有必要，且資料經過提供者處理後或經蒐集者依其揭露方式無從識別特定之當事人。」對此，亦得參照憲法法庭111年憲判字第13號判決主文第一項及相關法律上意見（第30至59段）所進行去識別化措施、匿名資料之探討，即要求資料經去識別化處理即符合該條款要求，而得蒐集、處理或利用，以利資料多元分享運用，詳見本文伍之一之（二）之2之（2）之A單元。

[30] 個資法第16條第5款：「公務機關對個人資料之利用，……，並與蒐集之特定目的相符。但有下列情形之一者，得為特定目的外之利用：……五、公務機關或學術研究機構基於公共利益為統計或學術研究而有必要，且資料經過提供者處理後或經蒐集者依其揭露方式無從識別特定之當事人。」

[31] 個資法第20條第1項第5款：「非公務機關對個人資料之利用，……，應於蒐集之特定目的必要範圍內為之。但有下列情形之一者，得為特定目的外之利用：……

定「資料經過提供者處理後或經蒐集者依其揭露方式無從識別特定之當事人」，則得為原始蒐集特定目的外之利用。歐盟一般資料保護規則（General Data Protection Regulation，下稱 GDPR）前言第 26 點（下稱 Recital 26）指出，GDPR 不適用於匿名化資訊（anonymous information），故不干涉匿名化資訊之處理（包括基於統計或研究目的）[32]。美國某些法規亦肯認去識別化資料的重要性與效用，其中，健康保險可攜性及責任法（The Health Insurance Portability and Accountability Act，下稱 HIPAA）[33]規定，受保護健康資料（protected health information，PHI）倘經去識別化將不受拘束[34]，不限制其運用或揭露[35]。

（三）小結

去識別化（匿名化）有助於資料蒐集者遵守個資保護義務，又能分享資料予他人或大眾；可在資料流通自由與隱私保護間取得平衡，在釋出有價值卻敏感資料時，仍能降低資料與個人之間連結可能性[36]。惟當前可供比對、連結資料來源多元，去識別化資料難以絕對排除被再識別化之可能。因此，去識別化若做得完善，促進資料供各種公、私運用時，尚可保護隱私；若做得不完善，則會侵害個人健康、尊嚴、聲譽或財務[37]。可知，資料去識別化的進行應審慎為之。

五、公務機關或學術研究機構基於公共利益為統計或學術研究而有必要，且資料經過提供者處理後或經蒐集者依其揭露方式無從識別特定之當事人。」

32 Tene, *supra* note 2, at 1239.

33 45 C.F.R. § 164.514(b).

34 GARFINKEL, *supra* note 5, at 4-5.

35 THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 6.

36 Brasher, *supra* note 10, at 217.

37 NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS (NCVHS), RECOMMENDATIONS ON DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION UNDER HIPAA 12 (2017), https://www.ncvhs.hhs.gov/wp-content/uploads/2013/12/2017-Ltr-Privacy-DeIdentification-Feb-23-Final-w-sig.pdf (last visited Apr. 14, 2022).

二、同意之取得與去識別化資料之利用

關於個資的去識別化、其後的利用應否以取得資料主體同意為必要，分述如下：

（一）同意為基礎之個人資料保護法制與侷限

1967 年 Alan F. Westin[38]即指出：「隱私乃個人……得決定其何種資料於何時、如何提供予他人之一種權利。」當前國際上，許多立法亦以經資料主體「同意」及協助該決策所需「告知」為保護個資之重要機制[39]。惟並非所有的資料蒐集、處理或利用均須以取得同意為前提，可能例外規定公益或執法考量而毋庸資料主體同意，即可運用個資，如傳染病之通報。此外，個資倘經去識別化、匿名化，不再視為個資，亦不再以取得資料主體同意為必要[40]。

倘經資料主體同意，則應以授權目的而蒐集、處理或利用個資。按歐盟 GDPR 第 5.1.（b）條，個資的蒐集必須基於特定、明示與正當目的（specified, explicit and legitimate purposes），且進一步處理的方式不得與原始目的不相容（incompatible）。再者，取得同意的適當性越欠缺，則越可能造成隱私侵害[41]。

惟對大型資料庫而言，倘欲事後取得同意，將遭遇實際挑戰，可能須承擔聯繫大量、甚至數以百萬計個人所衍生成本。如在醫療數年後才要聯絡，可能已搬家、死亡、不欲回憶不愉快或驚恐的心理創傷經驗、或違反資料主

---

38 ALAN F. WESTIN, PRIVACY AND FREEDOM 5 (1967)；另參司法院釋字第603號類似見解。

39 Fred H. Cate, *Protecting Privacy in Health Research: the Limits of Individual Choice*, 98 CAL. L. REV. 1765, 1766 (2010); 翁清坤（2013），〈告知後同意與消費者個人資料之保護〉，《臺北大學法學論叢》，87期，頁246。

40 EL EMAM & ARBUCKLE, *supra* note 25, at 2-3.

41 KHALED EL EMAM, GUIDE TO THE DE-IDENTIFICATION OF PERSONAL HEALTH INFORMATION 153 (2013).

體意願而洩漏予他人[42]。又如，難以想像 Google 須聯繫無數用戶，取得同意以先前搜尋關鍵字來預測流感？縱技術可行，也不會有業者願承擔此種成本[43]。其實，告知後同意成效不彰，除民眾多半不願費心閱讀隱私權政策外，大數據的資料再利用常不在最初蒐集目的內，如何能針對事前不可預測、解釋之資料運用方式、目的而預先進行通知[44]？

倘資料的運用皆須事前取得同意，許多科學突破將不會發生；因倘所能從事者皆事先取得積極授權使用之規模較小研究，幾乎難有創新發現[45]。另有證據顯示，當授予同意者或不同意者對於重要特徵有所差異，亦將導致資料集有所偏差；如都會、低學歷、低社經男性，較易同意參與實驗研究[46]。

（二）原始蒐集目的外的第二次利用未必以可識別個資為必要

如前揭，倘經資料主體同意，得以授權目的蒐集、處理或利用其個資。就健康醫療領域而言，可識別健康資料可供原始、主要授權目的（primary purpose）之用，如病人醫療照護之用。個人健康資料之第二次、衍生目的，則非供直接醫療照護之用，如分析、研究、品管、公衛、付款、行銷。如欲供超出原始目的範圍外之第二次目的（如公衛調查、研究）之用[47]，除符合法律例外規定或經去識別化外，應另取得同意。

當原始授權目的需要可識別個人的資料，去識別化非屬可行的選項，如醫療過程不可能隱藏病人身分。而利用健康資料供第二次目的之用，則病人常毋須可識別的，如許多健康服務研究、醫學生訓練或政策成效評估，不以

---

[42] EL EMAM & ARBUCKLE, *supra* note 25, at 2; *Id*. at 41; THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 28.

[43] Viktor Mayer-Schonberger、Kenneth Cukier（著），林俊宏（譯）（2013），《大數據》，頁216，天下文化。

[44] Dennis D. Hirsch, *The Glass House Effect*: *Big Data, the New Oil, and the Power of Analogy*, 66 ME. L. REV. 373, 390-91 (2014).

[45] Tene, *supra* note 2, at 1247.

[46] EL EMAM & ARBUCKLE, *supra* note 25 at 2; EL EMAM, *supra* note 41, at 44.

[47] EL EMAM, *supra* note 41, at 2.

能識別病人身分為必要。惟有時，去識別化資料庫中病人資料，有再識別化必要；如公衛研究發現病人曾曝露於某病毒（如近來肆虐之 Covid-19），乃須重行識別該人，以進行疫調追蹤[48]。

（三）去識別化資料之利用應否先取得同意為必要

1. 去識別化資料之利用不以取得同意為原則

如前揭，經去識別化資料不再被視為個資，不再以取得資料主體同意為必要，即可供原始蒐集目的外之利用或分享[49]。因此，當取得同意乃不切實際[50]、欠缺法令強制要求分享、蒐集個資前曾為匿名化之保證、資料蒐集者有裁量權而不欲分享個資、或非以可識別個資為必要時，則藉由去識別化之手段，即可任意蒐集、處理、利用或分享資料，供原始蒐集目的範圍外之第二次目的之用[51]。

2. 少數主張應立法授權得決定是否分享去識別化資料

資料一旦去識別化後，可提供予原本無權接觸之人，民眾常亦不甚關心該資料後續發展；惟此過於短視，因再識別化漸趨容易。學者[52]乃有主張，為保護健康資料，應立法授權病人（資料主體）除得決定是否分享其醫療資料外，尚得決定是否分享其去識別化之醫療資料。立法者可以賦予病人選擇

---

48 *Id*. at 2-3.

49 EL EMAM & ARBUCKLE, *supra* note 25, at 2; Benjamin T. Van Meter, *Demanding Trust in the Private Genetic Data Market*, 105 CORNELL L. REV. 1527, 1529-30 (2020).

50 當欲取得同意而接觸當事人將違反其意願而洩漏情況予他人、進行接觸將造成當事人精神或其他社會性侵害、資料檔案被研究之當事人已歿或失聯、樣本數規模龐大而難以取得所有人同意且負擔沉重時，則常未能取得告知後、知情同意（informed consent）。EL EMAM, *supra* note 41, at 41; ENGAGING PRIVACY AND INFORMATION TECHNOLOGY IN A DIGITAL AGE 221 (James Waldo, Herbert S. Lin & Lynette I. Millett eds., 2007).

51 Elizabeth R. Pike, *Securing Sequences: Ensuring Adequate Protections for Genetic Samples in the Age of Big Data*, 37 CARDOZO L. REV. 1977, 2016-17 (2016).

52 Christopher R. Smith, *Somebody's Watching Me: Protecting Patient Privacy in Prescription Health Information*, 36 VT. L. REV. 931, 981-88 (2012).

退出分享、限制何時與對誰分享資料、或允許不受限制地分享；並應建立追蹤系統，在非法再識別去識別化資料時，能偵測出來。惟亦有反對過度強化去識別化資料之保護，因擔憂將限制有益社會進步之可分享資料數量[53]，甚至導致「資料公地之悲劇」（tragedy of the data commons）（詳如本文肆之一之（二）之 2 部分）。

（四）個資的去識別化應否先取得同意

關於個資的去識別化應否先取得資料主體同意？見解分歧。

1. 肯定見解

按歐盟資料保護指令（Data Protection Directive 95/46/EC，下稱 DPD）[54]或 GDPR[55]規定，除法定例外情形外，對於個資之處理應經資料主體明確同意。由於資料的去識別化將涉及資料某些成分之毀損或消除，顯屬 DPD[56]或 GDPR[57]所定義之「處理」（processing）；故除取得同意外，否則，該「處理」乃 GDPR 所禁止之行為[58]。

2. 否定見解

在英國 *Regina v. Source Informatics Ltd.*案[59]，針對藥劑師將病人交付的處方箋資料去識別化後出售而供行銷之用，英國健康部主張「未經資料主體同意而將處方箋資料去識別化，乃背信」；上訴法院則推翻下級法院「支持健康部主張」的判決。健康部主張，資料的去識別化，乃某種形式的處理，需要資料主體明示同意。但上訴人 Source 主張，歐盟 DPD 不適用於「資料

---

53 N. Nina Zivanovic, *Medical Information As A Hot Commodity: The Need for Stronger Protection of Patient Health Information*, 19 INTELL. PROP. L. BULL. 183, 192-93 (2015).

54 Council Directive 95/46, art. 7(a), 1995 O.J. (L281) (EC).

55 Council Regulation 2016/679, art. 4(11), 2016 O.J. (L119) (EU).

56 Council Directive 95/46, art. 2(b), 1995 O.J. (L281) (EC).

57 Council Regulation 2016/679, art. 4(2), 2016 O.J. (L119) (EU).

58 Benjamin Charkow, *The Control over the De-Identification of Data*, 21 CARDOZO ARTS & ENT. L.J. 195, 222 (2003).

59 Regina v. Source Informatics Ltd., [2001] Q.B. 424 (Eng. C.A.).

的匿名化過程」，如同不適用於「匿名化資料的利用、揭露」。上訴法院認為，病人對於其資料無財產權，只要不會危害隱私，則無權限制資料利用，故對於病人資料進行去識別化不構成背信。據此，任何促使其成為匿名化的資料不再受保護，而「促使其成為匿名化的過程」毋庸取得資料主體同意。類似地，學者亦有主張[60]，某些法規（如美國金融服務現代化法［Gramm-Leach-Bliley Act］）認為「聚合統計資料（aggregate data）非可識別個資」，故只要未揭露用戶身分，則未禁止業者利用該資料；再者，該法未要求，個資的去識別化或蒐集、利用去識別化資料前須先通知資料主體。其實，社會運作實況，人們亦常將所經歷或耳聞事件，未取得同意，即將資料主體姓名或事件內容適度去識別化或隱匿而進行分享，因談論他人是非或相關資料常是言論自由的起點。

3. 民意分歧

曾有研究發現，對於未經同意而運用去識別化（匿名化）資料[61]，81%受訪者感到生氣。惟某些生化學者則認為，去識別化資料已無涉個人自主性[62]。尤其，去識別化資料堪稱健康照護拼圖的關鍵一片，應容許可供與原始蒐集目的（如病人照護）無關聯之第二次目的（如公衛）使用而分享之[63]。據此，縱有再識別化風險，惟去除明顯的識別符號而至少仍可提供某些程度隱私保護，乃另有研究指出「某些民眾仍顯現出以去識別化方式分享資料之可觀意願」[64]。

---

[60] Charkow, *supra* note 58, at 220-23.

[61] Leonard J. Kish & Eric J. Topol, *Unpatients: Why Patients Should Own Their Medical Data*, 33 NATURE BIOTECHNOLOGY 921, 921-24 (2015); Pike, *supra* note 51, at 2016; *Id*. at 221-24.

[62] Jorge L. Contreras, *Genetic Property*, 105 GEO. L. J. 1, 54 (2016); Bahrad A. Sokhansanj, *Beyond Protecting Genetic Privacy: Understanding Genetic Discrimination Through Its Disparate Impact on Racial Minorities*, 2 COLUM. J. RACE & L. 279, 303 (2012).

[63] EL EMAM & ARBUCKLE, *supra* note 25, at 1.

[64] Barbara J. Evans, *Barbarians at the Gate: Consumer-Driven Health Data Commons and the Transformation of Citizen Science*, 42 AM. J. L. & MED. 651, 676 (2016).

（五）去識別化造成資料流失而面臨如何兼顧效用之挑戰

資料之蒐集、處理或利用倘經資料主體同意者，則可保存較完整之個資內容；但倘去識別化者，則不免造成資料流失而降低資料效能與用途。大數據時代，資料集分析（尤其利用自動化演算法軟體）的價值，得藉由資料點（data points）之間連結關係而發見某些模式，而極大化；相反地，去識別化乃用以去連結（delink）上開本得以點點滴滴蒐集個資與身分間的資料點關係。因此，如前揭，乃衍生一大挑戰[65]：即如何確保資料去識別化能有效進行，而仍能保留資料效用？

（六）小結

「知情（告知後）同意」為個資保護的重要機制，但並非所有的資料蒐集、處理或利用均以取得同意為前提。惟倘經資料主體同意，應以授權目的運用個資。倘供原始目的外之再利用，除法律例外規定或經去識別化外，則應另經同意。

經去識別化之資料不再視為個資，不在隱私法規適用範圍，故不再以取得同意為必要，即可供原始蒐集目的外之利用或分享予原本無權接觸之人。惟有少數倡議，應立法授權資料主體得決定是否分享其去識別化資料；但有反對保護之強化，因將限制有益社會進步之可分享資料數量。

個資的去識別化應否先取得資料主體同意，亦見解分歧。有認為，資料之去識別化顯屬法定之「處理」，仍應取得同意。否定見解則主張，促使資料成為匿名化之過程，毋庸取得同意。

本文認為，促使資料成為去識別化、匿名化之過程及其後的資料利用，均毋庸取得資料主體同意，乃因考量經去識別化或匿名化資料一般已不再被

---

[65] Stalla-Bourdillon & Knight, *supra* note 4, at 285；憲法法庭111年憲判字第13號判決第56段亦指出，「或有主張將個人健保資料處理成為完全不具還原識別可能性之匿名資料再予利用，……匿名資料固非全然不具學術研究價值，但已喪失病歷、醫療、基因及健康檢查資料作為學術研究樣本時可擇定變因交互比對、建立相關性之特性者，將無從達成……所欲追求之特別重要公益目的。」

視為個資，且資料分享利用常為社會發展進步所不可或缺，故有如前揭資訊公開、開放資料之倡議，雖不免須忍受被再識別化風險，惟所採去識別化或匿名化技術須能確實降低再識別化風險至可接受的合理風險門檻、檻值之下，才可釋出資料進行分享利用。再者，為兼顧保護資料主體自主控制權，在進行資料去識別化或匿名化之前，如歐盟規定[66]或我國憲法法庭 111 年憲判字第 13 號判決[67]，應賦予選擇退出權。

三、去識別化與匿名化、假名化概念之異同

如前揭，去識別化、匿名化、假名化概念均早已深植法律政策，某些隱私法規即納入該等概念而視其為一種可符合法律要求的方法，乃放寬管制而可移作他用，以求在隱私保護與資料效用間達成最佳平衡。然而，按各立法例、學者主張可知，對於「去識別化」、「匿名化」內涵，眾說分歧，欠缺統一精確定義，有視為不同概念，亦有視為同義詞而交替使用。另外，假名化與去識別化、匿名化之概念異同亦有釐清必要。

（一）去識別化之定義

去識別化，乃藉由個資的操縱（manipulated）而使得其後更難從該資料再識別某一個人並預防資料風險的一種過程[68]；即去除識別資料並合理確保不能由殘留資料再連結至該個人。

1. 連結性之去除

去識別化之定義，首先普遍均著重於強調連結性（association）之去除、或直接、間接個人識別符號（personal identifiers）之去除。例如，Ira S.

---

[66] 歐盟GDPR第7(3)條規定，資料主體有權得隨時撤回其同意。

[67] 針對個資被強制蒐集、處理、利用，憲法法庭111年憲判字第13號判決主文第四項認定，個人健康保險資料提供原始蒐集目的外利用，欠缺當事人得請求停止利用之相關規定，乃違反憲法第22條保障人民資訊隱私權之意旨。

[68] Hintze, *supra* note 6, at 113; NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS, *supra* note 37, at 1.

Rubinstein 與 Woodrow Hartzog 指出，去識別化乃將識別資料（identifying data）與資料主體之間的連結性去除之過程[69]。Chris Achatz 與 Susan Hubbard 指出[70]，去識別化乃將直接與間接的個人識別符號與連結性由資料中去除之過程，但仍殘留再識別化之風險。美國國家標準暨技術研究院（National Institute of Standards and Technology，下稱 NIST）亦主張，去識別化乃由所蒐集、利用、存檔與分享之資料中去除個資的一種工具[71]。澳洲昆士蘭省健康資訊服務署（Health Informatics Services, State of Queensland，下稱 Queensland Health）指出，去識別化乃將識別資訊（identifying information）由資料集中去除，但仍有可能會被再識別化，例如，當識別資訊仍被保管而與該去識別化資料集相結合時[72]。

2. 殘留資料不再具有識別某一個人之合理可能性

去識別化之定義，除上開連結性或個人識別符號之去除外，既然去識別化資料只要保有任何效用則不免仍殘留再識別化之風險，乃有立法例[73]或主張要求尚應有「資料用以識別某一個人之合理可能性」的判別標準。例如，如前揭，關於受保護健康資料的去識別化標準，按美國聯邦部門式（sectoral）立法之一的 HIPAA 規定[74]，健康資料倘未能識別某一個人、且沒有合理基

---

69 Ira S. Rubinstein & Woodrow Hartzog, *Anonymization and Risk*, 91 WASH. L. REV. 703, 758 (2016).

70 Chris Achatz & Susan Hubbard, *Us vs. Eu Guidelines for De-Identification, Anonymization, and Pseudonymization*, 20 J. INTERNET L. 1, 7 (2017).

71 GARFINKEL, *supra* note 5, at 1.

72 Queensland Health, De-Identification and Anonymisation of Data Guideline 10, https://metrosouth.health.qld.gov.au/sites/default/files/content/de-identification-and-anonymisation-of-data-guideline.pdf (last visited Aug. 17, 2023).

73 惟歐洲理事會(Council of Europe, CoE)、歐盟DPD與GDPR、APEC個資規範未對「去識別化」明文界定。

74 45 C.F.R. § 164.514(a); W. Gregory Voss & Kimberly A. Houser, *Personal Data and the Gdpr: Providing A Competitive Advantage for U.S. Companies*, 56 AM. BUS. L.J. 287, 311-12 (2019).

礎（reasonable basis）而相信「該資料可能用以識別某一個人」，則非屬個人可識別健康資料。

當前美國少數州全面式（comprehensive）立法亦要求，去識別化資料須不能合理地連結至某一個人、或某一裝置（device）而可連結至該個人。例如，2020 年 11 月加州公投通過加州隱私權法（California Privacy Rights Act of 2020，下稱 CPRA），擴張及取代加州消費者隱私保護法(California Consumer Privacy Act of 2018，下稱 CCPA)之適用。CPRA 第 14（m）條（即民法第 1798.140（m）條）亦規定，去識別化（Deidentified），係指資料不能合理地（reasonably）用以推論資料或連結（linked）至某一特定消費者，倘持有該資料之業者：（A）採行確保該資料不能連結（associated）至某一消費者或家戶（household）之合理措施；（B）公開承諾以去識別化形式維持、利用該資料且不會嘗試再識別化該資料，除非再識別化乃僅出於判斷其去識別化過程是否符合此處規定之目的；與（C）以契約課予資料接收者應遵守此處要求。

科羅拉多州隱私法（Colorado Privacy Act）第 6-1-1303（11）條亦規定，去識別化資料（de-identified data），係指資料不能合理地（reasonably）用以推論資訊（information）或連結（linked）至某一已被識別或可識別個人、或某一裝置（device）而連結至該個人，倘持有該資料之控管者：（a）採行確保該資料不能連結（associated）至某一個人之合理措施；（b）公開承諾以去識別化形式維持、利用該資料且不會嘗試再識別化該資料；與（c）以契約課予資訊接收者應遵守此處要求。維吉尼亞州消費者資料保護法（Consumer Data Protection Act）第 59.1-575 條亦類似於上開加州及科羅拉多州隱私法規定。美國統一個人資料保護法（UPDPA）第 2（5）條規定，去識別化資料，係個資經修改（modified）而去除所有可直接識別資料並合理地確保某一對於資料主體資訊未有私人認識或特殊接觸之人不能由該檔案而連結至該主體。

美國聯邦貿易委員會（Federal Trade Commission，下稱 FTC）承認，雖不能完全徹底地去除揭露風險（disclosure risk），但倘有合理基礎而可相信

「某一特定檔案中的殘留資料不能用以識別某一個人」，則應被視為成功的去識別化。美國 NIST 另一報告亦主張，去識別化，乃藉由 PII 之去除或模糊化（obscured）（又稱「掩蓋、遮蔽」[mask]或「困惑化」[obfuscated]）而足以讓殘留資料（remaining information）不能識別某一個人、且沒有合理基礎而可相信「該資料可能用以識別某一個人」[75]。

類似地，澳洲隱私法（the Privacy Act）第 6（1）條規定，當資料不再涉及某一可識別（identifiable）個人或合理地可識別（reasonably identifiable）個人時，則該個人資料乃屬去識別化。因此，澳洲資訊監理辦公室（Office of the Australian Information Commissioner，下稱 OAIC）進一步指出，去識別化乃是一種過程，通常要求下列步驟而防止識別：（1）移除或變更個人識別符號（如名字、地址、日期或生日、及其他識別資料）；與（2）採行額外技術措施而能以某些方式地模糊化、聚合（aggregate）、變更（alter）或保護資料，以使任何個人不再能合理地被識別出來[76]。

---

[75] Achatz & Hubbard, *supra* note 70 at 7-8. 根據FTC見解，FTC隱私框架（privacy framework）僅適用於「合理地可連結」（reasonably linkable）至消費者之資料；而資料非「合理地可連結」，只有當業者（1）採取合理措施而確保資料乃經去識別化、（2）公開承諾不進行資料再識別、（3）以契約禁止後續的資料接收者嘗試再識別化。其中，針對第（1）項審查標準，FTC釐清：業者必須達到合理程度的正當確信「資料不能合理地用以推論或連結至某一特定消費者、電腦或其他裝置」。FEDERAL TRADE COMMISSION, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS iv (Mar. 2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf; ERIKA MCCALLISTER, TIM GRANCE & KAREN SCARFONE, GUIDE TO PROTECTING THE CONFIDENTIALITY OF PERSONALLY IDENTIFIABLE INFORMATION (PII) ES-3, 4-4, NIST, https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-122.pdf (last visited Apr. 14, 2022).

[76] OFFICE OF THE AUSTRALIAN INFORMATION COMMISSIONER, WHAT IS PERSONAL INFORMATION? 15 (May, 2017), https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/handling-personal-information?a=2832 (last visited Apr. 14, 2022).

（二）匿名化之定義

所謂「匿名資料」（anonymous data），有主張，乃直接與間接的個人識別符號被去除或操縱，一併以數學與技術方式確保，防止被再識別化[77]。惟對於匿名化方式是否不可逆、不可還原的而絕對不能有再識別化風險或可能性，則見解分歧，如下述。其實，匿名化之定義乃類似於前揭去識別化之定義，惟二者差異之處，乃去識別化方式未特別要求必須不可還原的；因此，二者有視為不同概念，但亦有視為同義詞而交替使用。

1. 不可逆、不可還原之見解

本見解主張，匿名化資料必須將識別資訊永久去除，且匿名化方式必須不可逆、不可還原的並應確保不可能被再識別化。

(1) 學說

Ira S. Rubinstein 與 Woodrow Hartzog 主張，匿名性（anonymity）或匿名化（anonymization）之概念乃隱含對於身分保護之保證[78]。Frederik Zuiderveen Borgesius、Jonathan Gray 與 Mireille van Eechoud 主張，匿名化，乃將資料集中的可資識別變數（identifying variables）去除或修改，使得資料主體不再可被識別出來[79]。ISO/TS 25237:2008(E)與 Chris Achatz、Susan Hubbard[80]均指出，個資之匿名化乃去識別化之子類別，亦即藉由將直接與間接的個人識別符號去除並落實技術安全維護措施，使得資料絕不會被再識別出來（never be re-identified）（即「零再識別化風險」）；其乃不同於「利用某一鑰匙（如代碼、假名、演算法）而可能再連結至個人」之去識別化資

---

[77] Kelsey Finch, *A Visual Guide to Practical Data De-identification*, FUTURE OF PRIVACY FORUM (Apr. 25, 2016), https://fpf.org/2016/04/25/a-visual-guide-to-practical-data-de-identification/ (last visited Apr. 14, 2022).

[78] Rubinstein & Hartzog, *supra* note 69, at 710.

[79] Borgesius, Gray, & Eechoud, *supra* note 22, at 2118.

[80] *ISO/TS 25237:2008*, ISO, https://www.iso.org/standard/42807.html (last visited Apr. 14, 2022); GARFINKEL, *supra* note 5, at 2.

料。因此，匿名化可界定為「一種採用於個資的技術，以促成不可還原的去識別化」[81]。

Mike Hintze 主張[82]，匿名化方式須不可還原的，且須去除任何已知或可預見的可能性（any known or foreseeable possibility）而可連結資料任何部分至該資料原始涉及的某一個人。

Maciej Gawronski 等主張[83]，匿名化乃一種不可還原的過程，資料一旦匿名化則不再有能識別當事人之風險。匿名化須確實執行，則匿名化將非用以保護個資而係用以摧毀個資。惟實際上，資訊部門人員常便宜行事，而有佯裝匿名化或輕忽特定資料可識別化（如獨特的保單序號）之傾向。

(2) 立法例

如前揭，按歐盟 GDPR Recital 26 規定，資料保護原則不適用於匿名資訊（anonymous information），亦即不涉及某一已識別或可識別自然人的資訊，或不涉及「個資以一種資料主體不是或不再可識別的（is not or no longer identifiable）方式而被進行匿名化」的資訊。據此，如同歐盟舊法 DPD 的標準，GDPR 亦可謂採「零再識別化風險標準」（zero re-identification risk standard），比上開美國 FTC 所採「合理程度的正當理由信賴標準」（reasonable level of justified confidence standard）更加嚴格[84]。

歐盟 DPD 第 29 條工作小組（下稱 WP29）指出，「匿名化資料」（Anonymised data）乃指「曾經指涉某一可識別的個人，但不再有可能識別」

---

[81] Achatz & Hubbard, *supra* note 70, at 7; Borgesius, Gray, & Eechoud, *supra* note 22, at 2118.

[82] MIKE HINTZE, VIEWING THE GDPR THROUGH A DE-IDENTIFICATION LENS: A TOOL FOR CLARIFICATION AND COMPLIANCE 3, https://fpf.org/wp-content/uploads/2016/11/M-Hintze-GDPR-Through-the-De-Identification-Lens-31-Oct-2016-002.pdf (last visited Apr. 14, 2022).

[83] Katarzyna Kloc et al., *CHAPTER 1 Basic Compliance*, *in* GUIDE TO THE GDPR 3, 22 (Maciej Gawronski ed., 2019) ; Aleksander P. Czarnowski et al., *CHAPTER 3 Security*, *in* GUIDE TO THE GDPR 185, 202-203 (Maciej Gawronski ed., 2019).

[84] Achatz & Hubbard, *supra* note 70, at 8.

之「匿名資料」[85]。WP29 亦指出，匿名化乃以不可還原地防止識別之方式處理個資；有三個指標可用以判斷「匿名化是否不可還原或永久刪除」，即是否仍可（1）單獨挑出（single out）某一個人、（2）連結（link）檔案至某一個人、及（3）推論（infer）出關於某一個人資訊；故倘答案是否定的，則資料可被視為「匿名的」。可知，GDPR 要求，資料集須經匿名化（而非僅去識別化），才能排除於規範範圍之外[86]；換言之，匿名資料乃處於個資的對立面，故匿名資料的處理（包括供統計、研究目的）則不在 GDPR 拘束之列[87]。

德國 2003 年頒布（非現行有效）的聯邦資料保護法（Federal Data Protection Act，BDSG）第 30a(2)條[88]規定，當資料以資料主體不再可能識別的方式而被匿名化時，則可供不同目的利用。

美國 NIST 主張，匿名化資料被界定為「曾為可識別資料但已被去識別化」與「用以再識別化之代碼或其他連結不復存在」[89]。

澳洲 Queensland Health 指出，匿名化乃將識別資訊永久去除，且未分開保留該識別資訊[90]。

2. 可逆、可還原之見解

---

85 *Opinion 4/2007 on the Concept of Personal Data* 21;吳全峰、許慧瑩（2018），〈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，頁50。

86 *Opinion 05/2014 on Anonymisation Techniques* 3, 8-12, 23; Achatz & Hubbard, *supra* note 70, at 8-9.

87 Council Regulation 2016/679, recital 26, 2016 O.J. (L119) (EU).

88 Bundesdatenschutzgesetz [BDSG] [Federal Data Protection Act], §30a, para. 2, "Personal data …may be processed or used for another purpose only if they have been rendered anonymous in such a way that it is no longer possible to trace them to a specific person," https://sherloc.unodc.org/cld/uploads/res/document/deu/1965/federal-data-protection-act_html/federal_data_protection_act.pdf (last visited Apr 14, 2022).

89 MCCALLISTER ET AL., *supra* note 75, at 4-5, 4-6.

90 Queensland Health, *supra* note 72, at 10, 20, 27.

本見解質疑，前揭「將匿名化界定為不可還原的，故再識別化之可能性等於零或趨近於零」之「絕對的匿名化」見解，在當前常有眾多資料來源可供比對的數位環境下，可能不切實際。因此，乃有將匿名化界定為，去除直接、間接識別符號之過程，以提供「資料非可識別的」合理保證，而供資料分享。匿名化乃機率問題，仍然有非常微小的再識別化機率、可能性；倘欲分享有用處的資料，則難以保證零風險，因為非常微小的風險乃分享資料所創造益處必須忍受的抵換代價[91]。

(1) 以合理審查標準界定匿名化而不以「絕對不可還原」為必要

A. 歐盟匿名化定義採合理審查標準，不要求「絕對不可還原」要件

「匿名化」定義與「個資」定義二者堪稱一體正反二面。其中，針對「個資」定義，歐盟乃採合理審查標準（reasonableness test）。按歐盟 GDPR Recital 26 規定，「為判斷某一自然人是否可識別的，必須考量資料控管者或任何其他人所有合理可能（reasonably likely）用以直接或間接識別（如挑選出）某自然人之所有手段。」歐盟舊法 DPD Recital 26 亦有類似規定。

循此，針對「匿名後的資料將不再屬於個資」之界定，歐盟亦採合理審查標準。歐盟 WP29 曾指出，上開「合理可能用以識別之手段」的審查標準亦用於評估「匿名化過程是否足夠完備」，即識別是否已成為合理地不可能。就 DPD 而言，匿名的資料乃指「不論資料控管者或任何其他人在考量所有合理可能（likely reasonably）用以識別該個人之手段而該當事人仍未能被識別出來」之任何涉及某一自然人之資料。其實，識別可能性將受到每一具體個案的特定情境脈絡與環境的直接影響。由於受到研究方式、工具與運算能力演進的影響，欲窮盡地完全列舉「何種情況識別不再可能的」，乃不可能或不太有實用性。然而，針對得用以識別資料集中主體的低成本技術方法與可公開近用的其他資料集日益增加，資料控管者應投入等比的匿名化努力與成本（即時間與資源）以取得平衡。尤其，隨著資通訊科技發展，再識別化

[91] ARBUCKLE & EL EMAM, *supra* note 5, at 8; EL EMAM & ARBUCKLE, *supra* note 25, at 14.

風險日增；法律規範亦須因應科技發展而調整。有效的匿名化應能防止某一個人由資料集被單獨挑出、連結及推論出來。一般而言，去除直接識別元素本身並不足以確保「不再可能識別資料主體」；而尚常須視匿名化資料所打算處理的情境脈絡與目的為何，而採行額外措施防止識別[92]。近來，歐盟資料保護委員會（European Data Protection Board，下稱 EDPB）在 COVID-19 指引中仍提出，「匿名化」乃指利用一系列技術而移除資料經合理努力即能連結至某一已被識別或可識別自然人之能力。此一合理測試（reasonability test）必須考量客觀面向（時間、技術方法）與每一個案所不同的情境脈絡要素（contextual elements）（現象的罕見性，包括人口密度、資料屬性與數量）。倘資料未能通過此測試，則仍未成功被匿名化，仍屬 GDPR 規範之列[93]。可知，Recital 26 所建立的測試標準，無疑採風險為基礎的取向（risk-based approach），以界定何種資料應構成個資；倘有識別的合理風險，應被視為個資[94]。

類似地，英國 ICO[95]指出，倘個資可以完全被匿名化，就不再是個資而不適用資料保護立法；匿名化，即在考量所有合理可能（reasonably likely）用以識別個人之手段，而不可能由該資料本身或結合其他資料而識別出某一個人。而所謂「合理可能」按具體情境而定，須考量技術與法律上識別的可

---

92 *Opinion 4/2007 on the Concept of Personal Data* 21; *Opinion 05/2014 on Anonymisation Techniques* 5-6, 8-9.

93 *Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak* 5 (Apr. 21, 2020).

94 Michèle Finck & Frank Pallas, *They Who Must Not Be Identified: Distinguishing Personal from Non-Personal Data Under the GDPR*, 10 INT'L DATA PRIVACY L. 11, 14 (2020).

95 THE INFORMATION COMMISSIONER'S OFFICE, BIG DATA, ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DATA PROTECTION 58, 97 (2017), https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf; THE INFORMATION COMMISSIONER'S OFFICE, CHAPTER 2: HOW DO WE ENSURE ANONYMISATION IS EFFECTIVE? 13, https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf (last visited Apr. 14, 2022); THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 6.

行性（包括成本、時間、當前與未來技術發展）。可知，個人能否被識別，ICO 亦採相對標準而進行判斷[96]。

學者亦指出，匿名化，指更動個資，而完全無法識別出特定自然人；或非投入不成比例的時間、費用與人力，否則，不能識別。所謂「不成比例」，須客觀進行個案判斷，主要以「資料儲存者對於去匿名化所表現出可能的興趣」與「因而帶來的經濟利益可能使得就算付出很高成本也很合理」為決定性因素；倘願為「去匿名化」投入大量資源，且不會被認為不成比例，則不能認為屬匿名資料[97]。

歐洲實務判決亦有類似於此處對於「匿名化」、「個資」定義採取相對的（而非絕對的）解釋取向，如歐盟法院 *Breyer v. Bundesrepublik Deutschland*[98]案。該案中，*Breyer* 造訪德國聯邦政府網站，網站記錄與儲存訪客動態 IP 位址以防止網路攻擊並可啟動刑事程序，動態 IP 位址並非涉及某一已被識別自然人的資料而是瀏覽紀錄，而只有 *Breyer* 所使用網路服務提供者（internet service provider，ISP）才擁有能識別出他所需的額外資料。*Breyer* 則質疑，儲存動態 IP 位址侵犯其權利。

對此爭端，提出意見協助審理的法務官（Advocate General，下稱 AG）指出，關於歐盟 DPD Recital 26「資料控管者或任何其他人合理可能使用手段」（means likely reasonably to be used by the controller or by any other person），「任何其他人」不應被視為「任何可能的第三人」（其實，不能絕對排除「有任何第三人能揭穿當事人的身分」），以及「合理可能使用手段」非指任何手段而是合理且非法律禁止的手段[99]。歐盟法院判決亦指出，倘資料主體之

---

[96] Finck & Pallas, *supra* note 94, at 14.

[97] Maria Cristina Caldarola、Joachim Schrey（著），趙彥清、黃俊凱（譯）（2020），《大數據與法律實務指南》，頁182-184，元照。

[98] Case C-582/14, Breyer v. Bundesrepublik Deutschland, 2016, para. 45-49; Finck & Pallas, *supra* note 94, at 17-18.

[99] *Opinion of Advocate General Campos Sanchez-Bordona, delivered on 12 May 2016*, at point 65, 68, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62014CC0582&from=en (last visited Nov. 30, 2022).

識別為法律所禁止或因需要投入不成比例時間、成本與人力而在實務上不可行，乃非合理可能用以識別資料主體之手段，則識別風險實質上顯得微不足道；然而，當線上媒介服務提供者（online media services provider）（網站經營者）擁有法律方法、手段（如因網路攻擊而啟動刑事程序強制揭露）而能結合網路服務提供者（ISP）所擁有關於資料主體之額外資料而能識別資料主體，則構成合理可能用以識別資料主體之手段，因此，動態 IP 位址乃被視為個資。可知，本案乃屬間接可識別性（indirect identifiability），即「能促使資料主體被識別出之所有資料，毋庸皆須歸一人持有，才能構成個人資料」；因此，動態 IP 位址的用戶可能在他人的協助下而被識別出來，而屬於個資[100]。

有評論進一步指出，關於個資定義，*Breyer* 案判決或 AG 意見乃採相對的（而非絕對的）解釋取向，並非理論上而需實質上可能藉由已知第三人合法協助而進行識別，而不欲膨脹「可識別性」概念至「資料何時才能終非個資」的充斥著法律不確定性地步。*Breyer* 案法院乃遠離、拋棄零風險的個資詮釋方式；相反地，WP29 前揭 05/2014「匿名化技術」意見所提及匿名化技術乃屬抽象的統計學，只有完全加以遵守且無人能再識別，資料才可能被視為匿名的。對於所持有資料、資料被處理的結果與情境脈絡，*Breyer* 案則要求進行具體的測試。*Breyer* 案確立二項明確的測試標準：（1）對於未被法律禁止識別的控管者而言，倘識別需要不成比例的時間、成本與人力努力，則識別風險實質上顯得微不足道；（2）縱非第一項測試標準之情形，即使識別風險實質上並非微不足道，惟當法律禁止該控管者或第三人協助進行識別時，則資料仍可能是匿名的[101]。

---

[100] EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, EUROPEAN COURT OF HUMAN RIGHTS, EUROPEAN DATA PROTECTION SUPERVISOR & COUNCIL OF EUROPE, HANDBOOK ON EUROPEAN DATA PROTECTION LAW: 2018 EDITION 91-92 (2018), https://data.europa.eu/doi/10.2811/343461 (last visited Apr. 14, 2022).

[101] Daniel Groos & Evert-Ben van Veen, *Anonymised Data and the Rule of Law*, 6 EUR. DATA PROTECTION L. REV. 498, 501-02 (2020), https://edpl.lexxion.eu/data/article/16563/pdf/edpl_2020_04-007.pdf (last visited Nov. 30, 2022).

近來英國法院[102]對於上開 *Breyer* 案抱持審慎取向而主張，歐盟法院的判決乃奠基於德國法律體系（尤其刑事訴訟）的特定事實面向，因此，不應認為僅因「一方當事人能藉由法律規定近用資料而能識別某一自然人」之事實，而即認為該程序係一種合理可能使用之手段[103]。

B. 自相矛盾的歐盟匿名化定義

針對「匿名化」的定義，歐盟立法模式出現自相矛盾。如前揭，歐盟法規（DPD、GDPR）規定「只有資料以資料主體不再可能被識別出的方式保存時，才可被視為匿名化」，對此，倘按上開 WP29「不可還原、永久刪除」的要求，實近乎一種不能承受任何風險的「不可能標準」（impossibility standard），無疑與「合理審查標準」[104]出現自相矛盾的緊張關係，而被質疑是否為一種可行、良好的規範[105]。究其因，乃出於歐盟 DPD 與 GDPR 規定文義不甚明確所致。具體上，一方面，二者雖以一種「資料主體不再可識別」的方式界定「匿名」，似採較嚴格的定義；另一方面，二者似亦支持風險為基礎的取向，以合理標準（reasonableness standard）而對於「可識別的」概念加以設限，亦即應考量以「合理可能用以識別某個人之所有手段」之程度為限[106]。可知，歐盟立法模式採「風險為基礎」的審查基準，卻又同時附加較嚴格的審查基準[107]，而自相矛盾。

WP29 指出[108]，匿名化流程的妥適性應就個案而定，不論所使用匿名化技術為何，須不可還原地防止再識別化。惟對此，有主張，為了達成資料的

---

102 Mircom International Content Management & Consulting Ltd. v. Virgin Media Ltd. [2019] EWHC 1827 (Ch), para. 27.

103 Finck & Pallas, *supra* note 94, at 18; Groos & Veen, *supra* note 101, at 501-02.

104 RUBINSTEIN, *supra* note 14, at 8-10; Finck & Pallas, *supra* note 94, at 15.

105 Groos & Veen, *supra* note 101, at 500-01.

106 Stalla-Bourdillon & Knight, *supra* note 4, at 298; PAUL VOIGT & AXEL VON DEM BUSSCHE, THE EU GENERAL DATA PROTECTION REGULATION (GDPR): A PRACTICAL GUIDE 40 (2017).

107 Raffi Teperdjian, *The Puzzle of Squaring Blockchain with the General Data Protection Regulation*, 60 JURIMETRICS J. 253, 313 (2020).

108 *Opinion 05/2014 on Anonymisation Techniques* 10.

匿名化，不應有「倘經合理努力（reasonable effort）即可用以再識別化資料主體之元素」仍殘留在資料之中[109]；故「不可還原」定義亦應將「合理可能」手段列為前提要件。而倘資料按此成功匿名化，則不再是個資而不受個資立法拘束。

乃亦有主張[110]，GDPR 實則納入 WP29 見解[111]，而應解釋為「個資的匿名化，應以考量合理可能使用之所有手段為限（即排除須投入不成比例之不合理手段），而使其不可還原地不可能識別資料主體」；此考量標準不僅須視個案情境而定，其結果亦隨時間而變遷。故為評估某資料集是否確實經匿名化，須隨時合理地考量再識別化風險。因此，GDPR 在界定匿名化屬性時，藉由要求應考量控管者或他人合理可能用以識別之所有手段而維持某程度彈性，此似符合「個資是一種連續、延伸狀態」的通念。因可識別性乃一易變的概念，隨不同時間的處理情境而異。

(2) 不可還原的匿名化定義方式乃不切實際

倘將「匿名化」界定為不可還原的，即等同要求再識別化之可能性、機率等於零或趨近於零[112]；其實，不可還原的匿名化乃極端困難，甚至是不可能的[113]。究其因，當前有各種自動化處理資料的方式與可能，絕對的匿名化只在很少情況才能達成[114]。即使聚合（aggregated）或聲稱匿名化（purportedly anonymized）資料，仍有可能被再識別化[115]。據此，有質疑，零風險標準不

---

[109] 再識別化風險的評估，得藉由資料性質、資料利用之情境脈絡、可資運用之再識別化技術與相關成本之角度，而考量再識別化所需之時間、努力與資源。EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS ET AL., *supra* note 100, at 93-94.

[110] JEF AUSLOOS, THE RIGHT TO ERASURE IN EU DATA PROTECTION LAW 141-142 (2020).

[111] *Opinion 05/2014 on Anonymisation Techniques* 8-9.

[112] RUBINSTEIN, *supra* note 14, at 8-10.

[113] Borgesius, Gray, & Eechoud, *supra* note 22, at 2130.

[114] Maria Cristina Caldarola、Joachim Schrey（著），趙彥清、黃俊凱（譯），前揭註 97，頁182-184。

[115] Borgesius, Gray, & Eechoud, *supra* note 22, at 2130.

僅不切實際，且與實際上風險從未能徹底排除之「風險為基礎的取向」有所矛盾[116]。

(3) 匿名化用語無異過度承諾

「匿名化」用語無異形同過度承諾、過度保證（over-promise），形塑近乎完美（near-perfection）的期待，人們因此陷入一切將安然無恙的錯覺之中。難怪美國一些深受矚目的再識別化案例發生時，媒體一再宣告「匿名化之死」，實因人們也緣木求魚一再期待「不可能」發生之事[117]。

有鑑於匿名化資料漸有被再識別化之可能，Woodrow Hartzog 與 Ira Rubinstein[118]即主張：較為可行者，乃將匿名化概念界定為「最小化風險之一種過程」（a process of minimizing risk），而非「保證安全之一種狀態」（a state of guaranteed safety）。

Woodrow Hartzog 與 Ira Rubinstein 並指出，「匿名化」並非是一直充滿爭議的概念。歷來普遍認為「資料只要經匿名化，則不會造成隱私侵害風險」，學者專家以往亦滿意於此。惟研究人員近來一再證明「被認定是匿名化的資料集之中，個人仍可被再識別出來」；完美匿名化（perfect anonymization）的概念不幸淪為一種迷思。「匿名化」似乎變得不可靠，再識別化個案吸引媒體目光，並成為「是否為保護個資有效工具」之正反爭辯焦點。尤其，所指稱的匿名化失靈似將政策討論引入失控局面。雖決策者（如 FTC、WP29）已注意到其極限，但對於安全釋出資料所需技術仍欠缺全面觀照，且隱私法規大部分仍僵化不變[119]。

由許多再識別化實例（如後述）可證，沒有完美的匿名化，故應著重於過程（process）而非結果（output）。Ira S. Rubinstein 與 Woodrow Hartzog 甚至主張，完美匿名化機制的追求，不應成為周延可行資料釋出政策的敵人；

---

116 RUBINSTEIN, *supra* note 14, at 10.

117 Rubinstein & Hartzog, *supra* note 69, at 750-51.

118 Woodrow Hartzog & Ira Rubinstein, *The Anonymization Debate Should Be About Risk, Not Perfection*, 60 COMM. OF THE ACM 22, 22-24 (2017).

119 Rubinstein & Hartzog, *supra* note 69, at 708-09.

幾乎所有使用「匿名化」詞彙以描述資料安全性，均是誤導的、有時甚至是欺罔的，故在資料釋出政策與過程中，「匿名化」的詞彙應被揚棄，而應著重於「過程與風險」之概念，才能協助消費者設定較佳、較正確可行的期待[120]。Paul Ohm[121]亦主張，我們需要一個新詞彙（如清理［scrub］），用以描繪隱私保護為基礎的資料運作，而單純意謂著「僅努力而非成功」（only effort, not success）。

（三）假名化與去識別化、匿名化之異同

1. 假名化定義

如前揭，去識別化可能涉及「假名化」（亦有譯為代碼化、擬匿名化）（pseudonymization）；所謂「假名化」，乃以某一資料值（value）（假名、代碼）取代真實資料或身分[122]，因此，除取得鑰匙（上開假名、代碼）比對外，資料主體不會被識別出來[123]。GDPR 第 4（5）條界定「假名化」為「個資處理之方式，而使該個資在未利用額外資訊時，則不能連結至某一特定資料主體，且倘該額外資訊已被分開存放並受拘束於技術及組織措施，以確保該個資未能連結至某一已識別或可識別的自然人。」美國加州、維吉尼亞州、科羅拉多州前揭隱私法規均有類似界定。美國 UPDPA 第 2（14）條則界定，假名化資料為「欠缺『可合理地連結至資料主體身分或得藉以與其進行個人化聯繫的』直接識別符號之個資」。美國 NIST 界定「假名化」為「將直接

---

120 *Id*. at 704-08.因此，某些學者喜歡使用去識別化（deidentification）與再識別化（reidentification）勝過於匿名化（anonymization）與去匿名化（deanonymization）之用語，因後者隱含對於身分保護之保證。Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1744 (2010).

121 *Id*. at 1744.

122 Rubinstein & Hartzog, *supra* note 69, at 759.

123 Czarnowski et al., *supra* note 83, at 202.

識別個人的名字或其他資料以假名取代的一種轉換，而假名化讓屬於某一個人的資料在不同檔案間可進行連結。[124]」

據此，「假名化」是一種可還原的過程，可逆、可還原性（reversibility）乃使其有別於匿名化之關鍵[125]。

2. 假名化技術

常出於資安考量，而將資料集內具有敏感性的資料欄位為假名化、代碼化處理。假名化常見技術，如（1）特定編碼、加密（encryption）而將資料編碼產生代碼，但仍得以安全金鑰而將代碼反組為原本資料值[126]；（2）標記化（tokenization）以非敏感資料值取代資料中敏感部分，如信用卡號“4111 111 111 1234”被取代為“4281 **** **** 2819”；（3）資料模糊化（data blurring）以近似值而讓資料變得失去意義，如一張臉孔模糊的肖像僅代表那裡有一個人，卻未能識別出是誰；（4）擾亂（scrambling）乃重新排列資料中的字元順序，如“obfuscate” 變成“tocbusafe”。惟將資料假名、代碼處理，未必能達匿名化效果，若真假名對照表或代換演算法遭破解，可能易遭逆向推敲[127]。

3. 假名化有時等同去識別化

有主張，倘符合前揭 FTC 三段審查標準，假名資料在某些情況下可被視為去識別化，但適用的規範為何，仍有歧見[128]。因此，乃有認為「假名化」係某種形式的去識別化，以「分開保管並受制技術安全拘束」的假名或人造

---

124 GARFINKEL, *supra* note 5, at 16.

125 Czarnowski et al., *supra* note 83, at 202.

126 項靖、陳曉慧、楊東謀、羅晉（2015），《開放資料及其對政府治理與個人隱私影響之研究》，頁88，國家發展委員會。

127 Brad Perry, *Pseudonymization, Anonymization & GDPR*, MEDIUM (Nov. 16, 2018), https://medium.com/@brperry/pseudonymization-anonymization-gdpr-3dc8405dd465 (last visited Apr. 14, 2022); 財團法人電信技術中心（2017），《「通傳事業去識別化技術與相關技術規範研究」補助研究報告》，頁40-41、103，財團法人電信技術中心。

128 Voss & Houser, *supra* note 74, at 311-12; 美國UPDPA第2(14)條指出，假名化資料不含去識別化資料。

識別符號（如數字）代替資料主體身分或識別符號[129]。在美國某些醫療領域，「去識別化」與「假名化」被等量齊觀[130]。

4. 假名化不等同匿名化而仍有利於資安

藉由假名化技術保護隱私，仍可由資料效用獲益。惟按歐盟 WP29 解釋，由於假名化資料繼續使某一個人能被挑選出（singled out）且能在不同資料集間進行連結，故假名化不等同匿名化；因假名化可能容許可識別性的存在，故仍屬個資法（如 GDPR）拘束範圍內[131]。對此，歐盟顯採「匿名化方式須不可還原」的較嚴格見解。然而，GDPR Recital 26 雖明示，假名化資料仍屬可識別某個人之資料；但倘須耗費可觀時間與金錢才能再識別的匿名化資料則非個資，顯有不同結論，而此「匿名化」定義乃採合理審查標準，不要求「絕對不可還原」之要件[132]。可知，GDPR Recital 26 規定與 WP29 此處解釋，有所矛盾。

假名化雖非屬匿名化的一種方式，其再識別化風險顯然高於匿名化[133]，但其能降低資料集與資料主體原始身分間的可連結性，仍屬一種有用的安全措施。惟為了脫離個資定義的拘束，資料須經適當的匿名化，而假名化仍有所不足[134]。

---

129 Rubinstein & Hartzog, *supra* note 69, at 759; Achatz & Hubbard, *supra* note 70, at 7.

130 GARFINKEL, *supra* note 5, at 2.

131 *Opinion 05/2014 on Anonymisation Techniques* 10.

132 Voss & Houser, *supra* note 74, at 322.

133 VOIGT & BUSSCHE, *supra* note 106, at 15.其實，歐盟GDPR規範涵蓋不同層級的去識別化，即匿名化與假名化，而有不同程度的再識別化風險與資訊價值，且根據其再識別化的相對風險而讓該等資料分類受到不同程度隱私要求的拘束。Brasher, *supra* note 10, at 252.

134 Voss & Houser, *supra* note 74, at 322;在日本，假名化與匿名化資料因去識別化程度差異而供不同用途，「日本個資法之相關規定可知依個資去識別化程度之高低，可分為『假名化加工資料』及『匿名化加工資料』，前者只要加工至該筆資料無從識別特定個人即可，未要求至無法回復原狀之程度。而後者則除須加工至無法識別特定個人外，尚須至依社會一般人之手段、技術無法將之回復原狀。後者因特定個人識別性低，因其利用而損害當事人之風險較小，故而可利用於特定目的外，包括行政機關得將匿名加工之個資檔案提供給民間利用於產業創新。而前者

假名化資料乃非匿名化資料，其技術措施亦不足以豁免 GDPR 之拘束；但 GDPR 仍認為，假名化足以降低風險並有益履行個資保護與安全維護義務[135]。GDPR 乃要求資料控管者應實施假名化等合理與適當措施，以藉由設計著手與預設保護個資（protect data by design and by default）[136]。因此，GDPR 無異提供讓資料假名化的重要誘因。

（四）小結

綜上，「去識別化」與「匿名化」之定義，均係指將可識別個人之符號、資訊去除、操縱而避免被再識別化。惟某些立法例、學者或文獻僅將「去識別化」描述為一種過程，而「匿名化」卻尚要求「須不可還原的」。相對地，有些則未區分「去識別化」與「匿名化」之不同，而將二個詞彙等量齊觀而交替使用；類似地，除在本文本單元（即貳之三）關於定義探討而介紹某些二者係不同概念之主張外，原則上，本文內文亦將「去識別化」與「匿名化」視為同義詞而交替使用，已如前揭。

另有認為「假名化」係某種形式的「去識別化」，故二者有時被等量齊觀[137]；「假名化」雖不符合 GDPR 匿名化個資保護應有要求，仍有助安全維護[138]。

如有其他資料得組合比對即可恢復原狀，故一旦提供給第三人，則被識別出特定個人之可能性高，故其利用目的被限於機關內之分析、研究用，以確實保護當事人之權益。」參憲法法庭111年憲判字第13號判決范姜真媺教授意見書，頁11；范姜真媺（2020），〈匿名加工資料制度之創設：因應大數據時代日本個人資料保護法之新進展〉，《東海大學法學研究》，59期，頁20、39-41。

135 *Pseudonymization according to the GDPR (Definitions and Examples)*, DATA PRIVACY MANAGER (Nov. 2, 2021), https://dataprivacymanager.net/pseudonymization-according-to-the-gdpr/ (last visited Apr. 14, 2022).

136 Achatz & Hubbard, *supra* note 70, at 1, 9.

137 GARFINKEL, *supra* note 5, at 2.

138 Achatz & Hubbard, *supra* note 70, at 9.

四、去識別化之方法

去識別化（匿名化）乃用以去除個資的工具，當資料欲供其他目的或外部人取用時，乃格外重要。去識別化之有效性，常須視具體個案情況而定。去識別化常非靠單一技術達成，而是不同方法取向、工具與演算法之集合而用於不同類型資料，進而產生不同程度的有效性。惟去識別化技術、方法越嚴謹，資料效用將越低[139]。其實，倘欲藉由去識別化、匿名化提供隱私保護的保證，只有匿名化技術經適當設計安排才能達成，亦即匿名化程序的前提要件（即情境脈絡）與目標必須加以釐清，才能達成所欲尋求實現之匿名化程度[140]。

針對去識別化之方法、程序，此處以發展相對成熟完整、具有全球影響力[141]的美國 HIPAA 規定為例加以說明，以供借鏡。按 HIPAA[142]規定，有「安全港」（safe harbor）與「專家認定」（expert determination）二種去識別化之方法、程序[143]。為符合 HIPAA 對於健康資料去識別化所要求之標準（亦即「無合理基礎而相信資料可用以識別某一個人」[144]），受拘束機構、主體（covered entity）（如多數健康照護提供者與機構）應符合下列二要件之一：（1）取得「專家認定」，即由某一符合資格的統計專家所為之正式認定；或（2）對於美國衛生及公共服務部（Department of Health and Human Services，

---

139 GARFINKEL, *supra* note 5, at 1.

140 *Opinion 05/2014 on Anonymisation Techniques* 23.

141 例如，HIPAA安全港標準已為加拿大研究者、政府與商業組織所採用，歐盟藥品管理局（European Medicines Agency，下稱EMA）亦欲引用為臨床試驗資料去識別化的最低標準，以利廣泛分享，EL EMAM & ARBUCKLE, *supra* note 25, at 7；其亦常受我國相關討論所引用，如憲法法庭111年憲判字第13號判決黃昭元大法官部分不同意見書第18段、及vTaiwan網站（2015），〈法務部「個人資料利用與去識別化」簡報〉，頁16，https://talk.vtaiwan.tw/t/topic/150（最後瀏覽日：11/30/2022）。

142 45 C.F.R. § 164.514(b).

143 Diana Warner, *Safe De-Identification of Big Data Is Critical to Health Care Organizations Must Find A Way to Strike A Balance As They Work Through the Challenges and Concerns*, 15 J. HEALTH CARE COMPLIANCE 63, 64 (2013).

144 45 C.F.R. § 164.514(a).

HHS）所指「安全港」條款中所列之識別符號加以去除、且受拘束機構不得確實知悉（actual knowledge）「殘留資料（remaining information）可能被用以識別個人」[145]，才能確認健康資料已經去識別化而不再屬於可識別個人之資料[146]。

不論「去識別化」以何種模式、方法達成的，HIPAA 隱私規則（Privacy Rule）即不限制去識別化健康資料的運用或揭露，因已不再是「受保護的健康資料」[147]。

受拘束機構得用以認定、決定「健康資料已被去識別化」之二種方法，分述如下：

（一）專家認定模式

按專家認定模式，即某一人員具有以一般認可的統計與科學原則、方法而進行資料去識別化之適當知識與經驗，藉以認定：「經去識別化資料可能被資料預定接收者（anticipated recipient）單獨或結合其他可合理取得資料而用以識別資料主體為誰，而其風險屬微小（little risk）」[148]。

1. 專家如何評估資料識別之風險

---

145 *Summary of the HIPAA Privacy Rule*, HHS, https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html (last visited Apr. 14, 2022)；加拿大亦有類似要求，再識別化風險須非常微小，才可將較詳細資料揭露。El Emam et al., *The Re-identification Risk of Canadians from Longitudinal Demographics*, 11 BMC MED. INFORMATICS AND DECISION MAKING 1, 4 (2011), https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-11-46 (last visited Apr. 14, 2022).

146 45 C.F.R. § 164.514(b).

147 THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 6.

148 45 C.F.R. § 164.514(b)(1)(**i**).

去識別化之進行，並無單獨統一方法，反而常須結合各種技術與政策程序[149]。因科技、社會條件、與得用以識別誰是某資料主體之可供利用資料狀況一直在變化，故專家認定應定期再進行評估[150]。

並無明確量化的可識別、再識別化風險程度，可被普遍視為符合所謂「非常微小」（very small），其乃按每一個案諸多具體因素而定[151]，於評估風險時應納入考量[152]。

專家得用以衡量某一特定資料集可識別、再識別化風險之原則，包括資料持續性出現機率之重現性（replicability）[153]、外部資料來源之可取得性（data source availability）[154]、當事人資料可被分辨出來程度之可分辨性（distinguishability）[155]、以及其重現性、可取得性與可分辨性越大而被識別的風險就越大之風險評估（assess risk）[156]。

---

[149] 45 C.F.R. § 164.514(b)(1)**(ii);** THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 12.

[150] ¶410 HIPAA Privacy Rules: Coverage and Scope, 2005 WL 4172319.

[151] Warner, *supra* note 143, at 64.

[152] THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 10-11.

[153] 如血糖水準檢驗結果常多變，人口統計特徵（如生日）則相對穩定。

[154] 化驗報告結果通常不會連同個人身分而被揭露於醫療環境之外，但人口統計特徵（如出生、死亡）較常出現在公開資料源。

[155] 曾有統計，病人出生年份、性別與三位數郵遞區號之資料結合，約僅對於0.04%美國居民，具有識別獨特性；但病人出生日期、性別與五位數郵遞區號之資料結合，對於超過半數美國人，將具有識別獨特性。

[156] 人口統計特徵具有高度可分辨性、高度重現性、與可在公開資料源中取得；化驗報告的資料值可能很具可分辨性，但罕具有重現性、罕可公開取得。THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 13; Warner, *supra* note 143, at 64.類似地，EMA亦指出某一變數（variable）能否被判定為「識別符號」（直接或準的，direct or quasi）的三項前提，包括重現性（replicability）、可分辨性（distinguishability）及可知悉性（knowability，即競爭對手[adversary]必須知悉資料主體相關的識別符號，以再識別化）。倘競爭對手對於變數一無所知，則難以進行再識別化之攻擊。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 48, https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-

2. 專家得用以去識別化之技術方法

類似於歐盟、英國 ICO、美國 NIST[157]等，美國衛生及公共服務部人權辦公室（DHHS Office for Civil Rights，下稱 OCR）亦提出得用以進行資料去識別化、匿名化而降低風險[158]（risk mitigation）之技術、方法如下：

（1）抑制（suppression），亦即在資料釋出之前，去除或排除某些特徵。倘檔案的某些重要內容深具風險性，則須進行某一項全部特徵的抑制（如郵遞區號特徵之去除）。倘某一檔案明確具有分辨性（如某特定地區大學的校長、某一郡之中僅有一位年收入超過 50 萬美金的個人），亦應進行抑制。

（2）概括化（generalization）（有時稱「簡化」[abbreviation]），亦即將資料轉化為更加抽象的呈現方式。如五位數郵遞區號概括化為四或三位數，以減少詳細（granularity）程度[159]。

---

data_en-3.pdf (last visited Nov. 30, 2022); Groos & Veen, *supra* note 101, at 501.

[157] 類似於此處HIPAA「受保護的健康資料」之去識別化、匿名化技術，各國或各國際組織均採用相去不遠而可適用於各種個資類型（包括健康等敏感性及一般個資）的去識別化、匿名化各種技術。例如，歐盟WP29於2014年所提出「匿名化技術」意見，分析各種技術的效用，包括：(1)雜訊加入（noise addition）：改變原始資料的精準度，如身高顯示正負10公分而非一個精確數字。(2)排列變更、擾動（permutation）：將某些欄位的資料彼此對調交換，如原屬A的資料移至B，而B的資料移至他人。(3)差分隱私（differential privacy）：釋出資料前，預先為適當的雜訊加入等處理，以供第三方資料查詢使用。(4)概括化（generalization）包括資料聚合（aggregation）與K匿名（k-anonymity）：即隱身於較大群體之中（hiding in the crowd）的概念，當每一個人隸屬於某一較大團體，則團體中的任何檔案均可對應至每一個人；因此，即將k個個人的資料以概括的群體方式聚合表示，以避免單一個人的資料被辨識出來；當K值愈高，對於個人隱私的保護程度也會愈高。例如，所在資料以一個國家代替某一城市。Perry, *supra* note 127; 項靖、陳曉慧、楊東謀、羅晉，前揭註126，頁77-81; *Opinion 05/2014 on Anonymisation Techniques* 12-17; THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 51-53; MCCALLISTER ET AL., *supra* note 75, at 4-6.可知，針對敏感性或一般個資的去識別化、匿名化技術並無太大差異，惟資料釋出的可接受風險門檻、檻值卻有所不同，詳如本文伍之一之（一）之3單元。

[158] THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 19; Brasher, *supra* note 10, at 214-16.

[159] Ohm, *supra* note 120, at 1714.

（3）排列變更、擾動（perturbation），亦即某些特定資料值（specific values）應以同樣特定但不同的資料值取而代之。如病人年齡可以真實年齡的 5 年區間內的隨機數字而記述之。類似地，所呈現郵遞區號乃在原始數字加減 3 之內。

上開任何一種方法的運用，不排除仍可同時運用其他方法（如概括化與抑制常一併運用）[160]，以減緩風險。由於每一受拘束機構與預定接收者各有不同，故所適用每一方法將各有優缺點，無某一特定方法可普遍適用，應由專家按具體個案[161]進行評估。

3. 其他領域的專家認定

由專家認定再識別化風險程度的做法，非健康醫療領域所獨有；在其他領域，亦由科學家、統計學家進行認定，以減緩風險。美國人口普查局（Bureau of the Census）、FTC 等政府機關亦指出，應採行如上開專家認定程序，以降低再識別化風險[162]。

臺灣亦有從事個資去識別化專家認定的驗證機構（如財團法人台灣商品檢測驗證中心），某些組織（如財團法人國家衛生研究院、中國醫藥大學附設醫院、國泰世華商業銀行股份有限公司等）已通過驗證[163]。

（二）安全港模式

第二種為「安全港」模式，即藉由去除某些識別符號而進行資料去識別化，且不得確實知悉將被用以再識別化。

---

160 THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 20.

161 Warner, *supra* note 143, at 64-72.

162 THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 10; Peter Swire & Jesse Woo, *Privacy and Cybersecurity Lessons at the Intersection of the Internet of Things and Police Body-Worn Cameras*, 96 N.C. L. REV. 1475, 1508-09 (2018).

163 財團法人臺灣商品檢測驗證中心，〈通過個人資料去識別化過程驗證客戶名單〉，https://www.bsmi.gov.tw/bsmiGIP/wSite/public/Data/f1685073214251.pdf（最後瀏覽日：08/17/2023）。

1. 去除 18 種識別符號

按 HIPAA 規定，應將下列 18 種識別符號從檔案中去除[164]，才符合去識別化要求：（1）姓名；（2）小於一個州的所有地理分區，包括街道地址、市、郡、區、郵遞區號（除非依人口普查局公開資料而取得郵遞區號前 3 個數字）；（3）直接連結至某一個人之所有構成「日期」之要素（年度除外），包括生日、入院日期、出院日期、死亡日期、超過 89 之所有年齡等；（4）電話號碼；（5）傳真號碼；（6）e-mail 帳號；（7）社會安全碼；（8）醫療檔案編號；（9）健康保險受益人編號；（10）帳號；（11）證照／駕照編號；（12）車輛識別特徵與序號（包括車牌號碼）；（13）裝置識別特徵與序號；（14）網址；（15）IP 位址；（16）生物識別特徵（包括指紋、聲紋）；（17）全臉照片影像與其他類似影像；及（18）其他獨特的識別號碼、特性或代碼[165]，但「隱私規則」另有規定得用以再識別則不在此限。

因此，HIPAA 藉由上開某些識別要素之去除，而允許去識別化資料之揭露、釋出。包括病人身高、體重、種族、出生年份、醫生名字等在內之資料，將不被視為可識別個人。故去識別化的病人檔案可能如下：（1）編號 24245 病人，高加索男性、6 尺高、1948 年生；Blue Cross/Blue Shield 承保，愛達荷州看 Jones 醫生，膽固醇而服用 Lipitor，胃食道逆流而服用 Nexium；（2）編號 33632 病人，1983 年生亞裔女性、5 尺 3 吋高、125 磅，華盛頓看 Smith 醫生，United Health Care 承保，去年春季由原廠避孕藥改為學名藥。檢驗後，季節性過敏而開始使用 Flonase[166]。

2. 不得「確實知悉」將被用以再識別化

---

164 45 C.F.R. § 164.514(b)(2)(i).

165 THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 7-8.

166 Jennifer L. Klocke, *Prescription Records for Sale: Privacy and Free Speech Issues Arising from the Sale of De-Identified Medical Data*, 44 IDAHO L. Rev. 511, 512 (2008).

按 HIPAA 規定，受拘束機構不得「確實知悉」上開去除 18 種識別符號的資料將被單獨或結合其他資料而用以識別該資料主體[167]。研究顯示[168]，以安全港模式進行去識別化之健康檔案，被再識別化風險將低於 0.25%。

OCR 曾解釋，所謂「確實知悉」，即「清楚且直接知悉」（clear and direct knowledge）該去識別化資料可能被單獨或結合其他資料而用以識別資料主體[169]。例如，（1）受拘束機構知悉：「檔案中列有病人職業（前州立大學校長）」，該資料幾乎只要結合其他額外資料（如年齡、居住的州名）即可清楚識別病人身分，因此，僅去除法律[170]臚列之識別符號是不夠的，除非亦去除病人職業資料，否則，難謂符合去識別化標準之「安全港模式」；（2）罕見臨床事件可能造成清楚且直接的識別。如去識別化資料提及某一病人曾生產罕見多胞胎並經媒體報導，受規範拘束之主體已注意到該報導，故必定知道該去識別化資料之主體可能被資料接收者識別[171]。據此，倘資料本質上有被再識別化風險、而受拘束機構清楚且直接知悉之，則該資料將不符合去識別化標準之「安全港模式」要求。

OCR 另明確指出[172]：「不得確實知悉」之要求，非指對於再識別化風險之研究、方法僅有「概括知悉」（general knowledge）「存在於理論上的可能性」（theoretically possible），而係指對於某一系爭特定資料集之再識別化風險有「特定、具體知悉」（specific knowledge）。故倘清楚且直接確實知悉「殘留資料可能用以再識別個人」，則非屬真正的去識別化[173]。惟此項

---

167 45 C.F.R. § 164.514(b)(2)(ii)；州法（如New Hampshire）亦有類似去識別化要素之規定，N.H. CODE ADMIN. R. He-C 1601.04.

168 Kathleen Benitez & Bradley Malin, *Evaluating Re-identification Risks with Respect to the HIPAA Privacy Rule*, 17 J. AM MED INFORM ASSOC. 169, 169-177 (2010), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3000773/pdf/jamia000026.pdf (last visited Apr. 14, 2022).

169 THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 27.

170 45 C.F.R. § 164.514(b)(2)(i).

171 THE DHHS OFFICE FOR CIVIL RIGHTS, *supra* note 8, at 27-28.

172 NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS, *supra* note 37, at 6.

173 EL EMAM & ARBUCKLE, *supra* note 25, at 7-8.

要求，卻未能充分被理解或契合當前運作實況，因越來越多後續利用涉及資料集的結合而劇烈改變去識別化時所做的風險評估，致漸難以落實上開要求[174]。

（三）專家認定與安全港模式之優劣、再識別化風險

許多受拘束機構雖承擔去識別化責任，內部卻無較具專業能力的統計專家，乃創造安全港模式的較簡便替代方案；惟安全港雖較普遍地被運用，卻是尚未標準化的模式，不同執行者可能產生歧異的運作結果[175]。安全港模式不論不同資料集之各自特性為何，均一體適用之。惟去識別化勢將降低資料品質與效能，其效果須按資料集與預定用途差異而進行判斷；不同去識別化方式，縱運用於相同資料集，仍會產生不同結果。

相對地，專家認定模式之優勢，可針對個別特定資料集之不同風險而量身打造去識別化方式；因其得針對資料集不同特性而運用一系列不同方法，所產生每一種不同的結果仍可被認為屬於去識別化。縱專家認定模式較具上開優勢，但比起安全港模式，專家認定模式卻較少被使用[176]，因仍須確保識別風險「非常微小」[177]，且需較多諮詢過程而更耗費成本，市場上可供聘用專家人數亦過少。近來乃有呼籲，針對專家認定模式之使用，應建立相關準則與標準，包括專家可使用方法、達成結果、可擔任專家之能力與資格最低標準應透明化[178]。

此外，縱善加運用上開安全港與專家認定任一模式而去識別化，難免仍有識別、再識別化風險。風險縱非常微小，仍非是「零」[179]；只要個資衍生

---

174 NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS, *supra* note 37, at 6.

175 CENTER FOR DEMOCRACY & TECHNOLOGY (CDT), ENCOURAGING THE USE OF, AND RETHINKING PROTECTIONS FOR DE-IDENTIFIED (AND "ANONYMIZED") HEALTH DATA 6, 9 (2009), https://cdt.org/wp-content/uploads/pdfs/20090625_deidentify.pdf (last visited Apr. 14, 2022).

176 NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS, *supra* note 37, at 8.

177 Warner, *supra* note 143, at 64-72.

178 NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS, *supra* note 37, at 8.

179 Deven McGraw & Alice Leiter, *Risk-Based Regulation of Clinical Health Data*

而來的資料留有任何效用，仍可能據以連結回去資料原始主人，而喪失去識別化本欲提供之隱私保護；故在決定如何、是否進行資料去識別化時，應一併考量該資料將如何被利用、分享或釋出，因再識別化風險可能性不易估算[180]。

（四）有限資料集

除包括受保護健康資料與去識別化資料外，HIPAA 隱私規則尚承認第三種類型資料，即未完全去識別化之「有限資料集」（limited data set）。因有諸多抱怨「去識別化安全港標準過於嚴苛」，2002 年 HIPAA 乃制定新條文[181]，而容許受拘束機構毋庸取得病人同意即可利用或揭露有限資料集而供研究、公衛與健康照護營運目的[182]之用，但須以該機構與資料集接收者有締結資料使用契約（data use agreement[183]）而受拘束（如非經契約或法律許可而不得使用或進一步揭露資料、採行適當安全維護措施、不得再識別資料或接觸當事人[184]）為前提[185]。

有限資料集乃去除多數類型的可直接識別資料[186]，而仍保留公衛、健康研究通常所需資料（如生日、治療日期與某些地理資料），故僅為部分去識

Analytics, 12 COLO. TECH. L.J. 427, 443 (2014).

180 GARFINKEL, *supra* note 5, at 1.

181 45 C.F.R. § 164.514(e).

182 45 C.F.R. § 164.514(e)(3)(i).

183 45 C.F.R. § 164.514(e)(4)(i).

184 45 C.F.R. § 164.514(e)(4)(ii).

185 Stacey A. Tovino, *The Use and Disclosure of Protected Health Information for Research Under the Hipaa Privacy Rule: Unrealized Patient Autonomy and Burdensome Government Regulation*, 49 S. D. L. REV. 447, 457 (2004); Deven McGraw & Alice Leiter, *A Policy and Technology Framework for Using Clinical Data to Improve Quality*, 12 HOUS. J. HEALTH L. & POL'Y 137, 144-45 (2012).

186 包括(1)姓名；(2)鎮或市、州、郵遞區號以外之郵遞地址；(3)電話號碼；(4)傳真號碼；(5) e-mail帳號；(6)社會安全碼；(7)醫療檔案編號；(8)健康計畫受益人編號；(9)帳號；(10)證照／駕照編號；(11)車輛識別特徵與序號（包括車牌號碼）；(12)裝置識別特徵與序號；(13)網址；(14) IP位址；(15)生物識別特徵（包括指紋、聲紋）；(16)全臉照片影像與其他類似影像。45 C.F.R. § 164.514(e)(2).

別化資料集[187]。例如，生於 2004 年 1 月 1 日、住德州休士頓郵遞區號 77002 小孩的哮喘病況資料可留在有限資料集中，但按去識別化安全港規定，同樣資料將不符合去識別化資料[188]。惟研究指出[189]，有限資料集的再識別化風險範圍從 10%至 60%，亦證實「有限資料集乃非經適當去識別化」。故有別於完全去識別化資料可供任何目的使用，有限資料集保留較多資料元素，用途乃有更多限制[190]。

（五）小結

綜上，去識別化之有效性，常須視具體個案而定。去識別化常並非靠單一技術，而是不同方法、工具與演算法之集合而產生不同程度的有效性。惟去識別化方法越嚴謹，資料效用將越低。

按美國 HIPAA，有安全港與專家認定二種去識別化之方法、程序。不論去識別化以何種方法達成，HIPAA 隱私規則即不限制去識別化健康資料的運用或揭露，因已不再是「受保護的健康資料」。惟縱善用上開任一方法而進行去識別化，仍難免有再識別化風險；風險縱非常微小，仍非是「零」。

參、去識別化資料之再識別化風險

歷來有主張，資料能安全永續分享的關鍵，在於去識別化（匿名化）；只要去識別化，即可自由利用、分享。此種觀念深植法律與政策之中，如前揭 HIPAA、歐盟等規範[191]。以去識別化兼顧個資保護與運用，理論上似可運作良好；惟實際運作上，難以萬無一失的方法而防止資料被再識別化，卻

---

187 GARFINKEL, *supra* note 5, at 26.

188 Tovino, *supra* note 185, at 458.

189 GARFINKEL, *supra* note 5, at 25-26; Benitez & Malin, *supra* note 168, at 169.

190 CENTER FOR DEMOCRACY & TECHNOLOGY, *supra* note 175, at 6.

191 Hartzog & Rubinstein, *supra* note 118, at 22.

又全然未犧牲效用。乃有認為，資料不是徹底去識別化，就是徹底有效用，二者不易兼顧[192]。

所謂「再識別化」，係指嘗試從已被去識別化資料中重新識別出身分之過程[193]，或指分析資料或與其他資料相結合而讓個人變得可識別之過程，有時稱「去匿名化」[194]。過去數年，藉由去識別化（匿名化）資料與包含有識別符號相關資料之間的交叉比對，而能再識別出當事人身分的一些惡名昭彰案例，更讓人懷疑去識別化仍能否有效保護當事人免於被追蹤、剖析[195]。如何確保資料分享予第三人運用而仍能維持去識別化、匿名化之有效運作，乃當前大挑戰[196]。

一、去識別化之功效漸受質疑

曾有雜誌刊登知名漫畫[197]：「在網路上，沒人知道你是條狗」。大數據、AI 時代之前，那可能是真的，因欠缺不同資料來源且交叉比對資料很繁瑣。當大數據能由更多資料、更快速電腦與更精進分析技術而運作，尤其 AI 能由看似不相干大量資料比對推衍關係，將擅長再識別化而威脅隱私，去識別化、匿名化概念漸成鏡花水月，惟當前法律運作卻繼續奠基此幻象之上。乃有建議應通盤評估隱私規範，卻猶未改善[198]。

---

192 Tene, *supra* note 2, at 1240.

193 GARFINKEL, *supra* note 5, at 9.

194 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 49.

195 Ira S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 INT'L DATA PRIVACY L. 74, 78 (2013), https://academic.oup.com/idpl/article/3/2/74/709082 (last visited Apr. 14, 2022).

196 Stalla-Bourdillon & Knight, *supra* note 4, at 285.

197 Peter Steiner, *On the Internet, Nobody Knows You're a Dog*, THE NEW YORKER (July 5, 1993), https://smg.media.mit.edu/library/steiner1993.html (last visited Apr. 14, 2022).

198 Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J. L. & TECH. 106, 128-29 (2019).

白宮報告[199]指出，當資料單獨存在且未能連結至某一特定人，匿名化是可行的。不幸地，大數據可運用許多技術而輕易讓匿名化落空，匿名化有時甚至賦予錯誤的隱私期待。在大數據環境中，去識別化、匿名化已漸失效，因已被去識別化、匿名化之資料仍可能與散落其他來源之資料相結合，而難以絕對確定「一個人不可能由某一資料集被再識別出來」。

白宮另一報告[200]指出，資料融合（data fusion）等技術使大數據分析更具威力，日益嚴厲挑戰隱私期待。當資料連結至某一當事人，雖可以科技除去連結性而匿名化或去識別化，但同樣亦得由同等有效的再識別化技術而將諸多支離破碎的片段再組合回去。類似地，藉由散落各處、各種不同資料（如交易、人口統計、甚至閱讀習慣）的彙整、統合所形成的數位軌跡（digital trails）[201]，而形成所謂「馬賽克效應」（mosaic effect）[202]。「馬賽克效應」與「再識別化」乃攀生兄弟[203]，因在易於進行各種不同資料彙整的大數據環境中，本來不具識別性之去識別化、匿名化資料可與其他大量、多元資料互相結合、比對時，可資識別特定個人的資料得萃取、推衍、再識別化出來，進而獲知該個人是誰、身分與喜好為何之相關描繪、剖析（profiling）[204]。

歐盟 WP29 指出，匿名化可能是可以保有益處又能減緩風險的一個好策略。一旦資料集被真正地匿名化，個人將不再是可識別的，個資法將不再

---

199 President's Council of Advisors on Science and Technology, *supra* note 9, at 38-39, 59, 97.

200 EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 8 (2014), http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf (last visited Apr. 14, 2022).

201 Jane B. Baron, *Property as Control: The Case of Information*, 18 MICH. TELECOMM. & TECH. L. REV. 367, 378 (2012).

202 EXECUTIVE OFFICE OF THE PRESIDENT, *supra* note 200, at 8.

203 Jesse W. Woo, *Smart Cities Pose Privacy Risks and Other Problems, but That Doesn't Mean We Shouldn't Build Them*, 85 UMKC L. REV. 953, 961 (2017).

204 EXECUTIVE OFFICE OF THE PRESIDENT, *supra* note 200, at 8, 59, 97; McKay Cunningham, *Privacy Law That Does Not Protect Privacy, Forgetting the Right to Be Forgotten*, 65 BUFF. L. REV. 495, 537 (2017)；翁清坤（2020），〈大數據對於個人資料保護之挑戰與因應之道〉，《東吳法律學報》，31卷3期，頁96。

適用。然而，由某些公開的案例與研究清楚顯示「欲由含有豐富個資的資料集中，建立一個真正匿名的資料集，非易事」，因工作所需而常仍須殘留相當的資訊。因此，被視為匿名化的資料集，仍可能與其他資料集相連結、結合，而導致某一或某些個人可以被識別出來[205]。就可進行連結之其他資料而言，英國 ICO 進一步指出，由於不易精準確認何種其他資料已可供取用或未來會釋出，因此，根本難以預測「以此種資料連結而再識別化之風險」為何[206]；惟隨著資通訊科技發展，再識別化之風險與日俱增[207]。2018 年取代歐盟 WP29 之 EDPB 亦指出，由於可用科技方法與再識別化技術的進展，可能難以達成與維持個資的匿名化狀態[208]。

Viktor Mayer-Schönberger 與 Kenneth Cukier[209]亦有類似發現，只要刪去個人識別符號後的去識別化、匿名化資料，不會損害任何人隱私，便能共享與分析，但那只有在小數據（small data）世界才如此；在大數據世界，資料數量與種類皆增加，要再識別身分，並非難事，只要給予足夠資料，不論多麼努力小心，都不易達成完美匿名化。

Frederik Zuiderveen Borgesius、Jonathan Gray 與 Mireille van Eechoud[210] 亦指出，法律雖區分個資與匿名化資料之不同；但就電腦科學而言，二者僅程度差異，而非屬性上全然涇渭分明。其實，不可還原之匿名化（irreversible anonymization）是不容易的，甚至是不可能的。資料是否充分匿名化，常難以事前評估；當有更多資料集可資運用，而促成拼圖式的身分識別（jigsaw

---

205 *Opinion 05/2014 on Anonymisation Techniques* 5.

206 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 18.

207 *Opinion 05/2014 on Anonymisation Techniques* 9.

208 European Data Protection Board, *EDPB Document on Response to the Request from the European Commission for Clarifications on the Consistent Application of the GDPR, Focusing on Health Research*, para. 47 (adopted Feb. 2, 2021), https://edpb.europa.eu/sites/default/files/files/file1/edpb_replyec_questionnaireresearch_final.pdf (last visited Nov. 5, 2022).

209 Viktor Mayer-Schonberger、Kenneth Cukier（著），林俊宏（譯），前揭註43，頁 217-219; Peter Segrist, *How the Rise of Big Data and Predictive Analytics Are Changing the Attorney's Duty of Competence*, 16 N.C. J. L. & TECH. 527, 572 (2015).

210 Borgesius, Gray, & Eechoud, *supra* note 22, at 2121.

identification），更是如此。當公務機關釋出越多資料（data），越可能進行資料結合而交叉比對，創造得以識別個人之資訊（information）。

國內亦有類似見解，如憲法法庭 111 年憲判字第 13 號判決黃昭元大法官部分不同意見書第 16 段指出，「按現行科學及技術條件下的匿名化技術，已越來越難以達成完全的匿名化（指徹底的去連結，而無從再還原識別當事人）。因此，所謂去識別化其實比較像是個光譜，有些去識別化方式仍容易還原連結，有些去識別方式則需耗費較多成本方能識別，但最終仍能識別。」

綜上，大數據可能使得去識別化、匿名化越來越難以有效運作。現在有更大型資料集與更有用的分析方法可資運用，去識別化、匿名化的資料與某一特定當事人的身分之間，有時僅有幾步之遙[211]。過去實例一再證明[212]，當再識別化技術越來越可行，去識別化、匿名化將會越來越無效[213]。

二、再識別化之實例與比例

1990 年代開始，美國發生一些知名事件（如下述 Weld、Netflix、AOL）[214]，顯示「明顯經去識別化的資料集仍易遭再識別化攻擊」；尤其，大數據帶動資料不斷誕生、蒐集與分析，及伴隨著技術演進與演算法進步，再識別化能力逐漸強化[215]。許多法律、技術領域學者乃認為，按此趨勢，可識別資料與非可識別資料間原本具有意義的分野恐難持續長久，進而質疑去識別化的有效性[216]；但去識別化的擁護者則指出「在為數眾多檔案中，僅有少量被

---

211 John Pavolotsky, *Privacy in the Age of Big Data*, 69 BUS. LAW. 217, 221 (November, 2013).

212 Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 257 (2013).

213 Segrist, *supra* note 209, at 571.

214 Ohm, *supra* note 120, at 1717-23.

215 Tene, *supra* note 2, at 1239.

216 RUBINSTEIN, *supra* note 14, at 3.

再識別出來」[217]，去識別化仍屬有效機制。針對當前不乏已明顯去識別化資料而仍被再識別化的例子[218]，略述如下：

例一，在去除可直接識別符號而仍保留人口統計資料（郵遞區號、生日與性別）與敏感性健康資料後，州保險相關機關有義務將某些醫療檔案供民眾研究之用；在 1997 年，學者 Latanya Sweeney 取得去識別化醫療檔案後，與可公開取得的選民註冊資料（包含相似的人口統計資料）相比對，進而再識別出麻州州長 William Weld 的醫療資料[219]。Latanya Sweeny 甚至指出，只要郵遞區號、生日與性別等三項無害資料，即可識別出 87%美國人民[220]。

例二，研究人員藉由分析 Netflix 電影評分（movie ratings）之匿名化資料集，再結合公告在其他網站之電影評分資料庫（Internet Movie Database，IMDb），即可判斷出某些參與評分者身分。研究人員另發現，倘競爭對手（adversary）知悉 Netflix 用戶某一期間所租電影，即可逆向工程發現所有觀看紀錄[221]。

例三，2006 年搜尋引擎業者 America Online（AOL）釋出 658,000 名用戶顯已去識別化（匿名化）的搜尋關鍵字資料，紐約時報記者卻能連結搜尋關鍵字至特定用戶而進行聯繫；如藉由所搜尋關鍵字（如仲介、園丁、健康、約會網站、寵物行為），用戶代號 4417749 能連結至一位 62 歲老太太[222]。

---

217 Polonetsky, Tene & Finch, *supra* note 8, at 600.

218 Tene, *supra* note 2, at 1239.

219 Rubinstein & Hartzog, *supra* note 69, at 711.

220 Tene, *supra* note 2, at 1240.

221 Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, *in* 2008 IEEE SYMPOSIUM ON SECURITY AND PRIVACY 111, 123 (The IEEE Computer Society & The Institute of Electrical and Electronics Engineers eds., 2008), https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4531148 (last visited Apr. 14, 2022); EL EMAM, *supra* note 41, at 6.

222 Viktor Mayer-Schonberger、Kenneth Cukier（著），林俊宏（譯），前揭註43，頁 218-219; Sheri B. Pan, *Get to Know Me: Protecting Privacy* and *Autonomy Under Big Data's Penetrating Gaze*, 30 HARV. J.L. & TECH. 239, 256 (Fall, 2016); Brasher, *supra* note 10, at 229.

故搜尋資料本身可用於識別個人；尤當以自己名字進行虛榮心搜尋（vanity search）而讓名字成檔案一部分，問題益形惡化[223]。

除上開知名三例外，各國尚有不少再識別化案例如下：

例四，分析歐洲某國 15 個月 150 萬手機用戶資料後，研究人員發現，只要根據某年某一小時四個場合資料，即足以確認某一手機基地臺數百碼內之某一個人位置，而可識別出資料集中 95%用戶，故非常難以保持匿名性[224]。

例五，縱未釋出原始資料（raw data），利用外部正確資料仍可去匿名化。二戰時，美國人口普查局提供街區等級資料縱未能識別出特定家庭，仍有助定位與遞送日裔美國人至集中營[225]。另利用可公開取得之 2010 年人口普查資料，研究人員可重建個人檔案（如年齡、種族、性別），並藉由與商業資料比對而以名字再識別出 17%民眾或 5,200 萬人[226]。

例六，蘇黎世大學研究者藉由連結可公開取得資料庫，再識別化瑞士聯邦法院 84%匿名化案件[227]。

例七，墨爾本大學研究者由本應匿名的醫療帳單檔案，再識別化病人。紐約計程車資料集被揭露後，藉由信用卡元資料（metadata）能識別出名人乘車地點、車資及小費[228]。

---

223 Anne Klinefelter, *When to Research Is to Reveal: The Growing Threat to Attorney and Client Confidentiality from Online Tracking*, 16 VA. J.L. & TECH. 1, 2 (2011).

224 Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent*, 93 TEX. L. REV. 85, 131 (2014);Pavolotsky, *supra* note 211, at 221.

225 Brasher, *supra* note 10, at 227.

226 Cara Brumfield & Jae June Lee, *The Risks and Rewards of Conducting A Census in the Digital Age*, 4 GEO. L. TECH. REV. 415, 423-24 (2020).

227 Simon Chandler, *Researchers Use Big Data And AI To Remove Legal Confidentiality*, FORBES (Sep. 4, 2019, 10:06 AM), https://www.forbes.com/sites/simonchandler/2019/09/04/researchers-use-big-data-and-ai-to-remove-legal-confidentiality/?sh=36b67f8f15f6 (last visited Apr. 14, 2022)

228 Manheim & Kaplan, *supra* note 198, at 128-29.

三、再識別化之技術、動機與風險評估

（一）再識別化攻擊

資料的去識別化、匿名化不再能堅不可摧地保護資料主體隱私，因有再識別化之可能性，即除可分析資料本身外，尚可連結去識別化資料至額外輔助資料而能重新識別[229]。去識別化之重要目標，乃防止未經授權之再識別化攻擊（re-identification attacks）[230]。

再識別化常發生於競爭對手以額外資料連結去識別化資料而再識別化某一資料集中個人之嘗試，亦即「連結攻擊」（linkage attack）。「攻擊」之用語乃由資安文獻借用而來，因此，進行攻擊之人被稱為「競爭對手」。而藉以識別資料集中個人之額外資料、資訊，則被稱為「外部、輔助、背景資料、資訊」（outside, auxiliary, background information）[231]。許多案例已證明，匿名化資料經與可取得額外資料交叉比對，可識別出本已匿名化之個人[232]。

當前大數據環境，不論自願或非自願揭露[233]，可供取用資料漸多，「連結攻擊」漸易。資料可供取用性強化了將非個資轉化為個資的能力，某些個資法規一一列舉僵化的識別符號，以明確界定個資範圍，但被批評[234]為恣意妄為的分類方式。如 HIPAA 列舉 18 種識別符號，無異假定「健康檔案中的其他資料不會用以再識別」，實屬謬誤而應修法改正。

---

229 Brasher, *supra* note 10, at 225-42; THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 49; EUROPEAN MEDICINES AGENCY, *supra* note 156, at 10.

230 GARFINKEL, *supra* note 5, at 9.

231 Rubinstein & Hartzog, *supra* note 69, at 734; Brasher, *supra* note 10, at 226-227; *Opinion 05/2014 on Anonymisation Techniques* 11-12.

232 Michelle M. Christovich, *Why Should We Care What Fitbit Shares?: A Proposed Statutory Solution to Protect Sensitive Personal Fitness Information*, 38 HASTINGS COMM. & ENT. L.J. 91, 112 (2016).

233 EL EMAM, *supra* note 41, at 137-138.

234 Brasher, *supra* note 10, at 227-28, 237.

（二）以殘留資料再識別化

如前揭的諸多案例顯示，經匿名化資料集對於資料主體可能仍會構成殘留風險（residual risk）。即使不再能精準地回復某一個人的資料檔案，但藉由可供取用（不論是否公開）的其他來源資料的協助下，仍可能點點滴滴蒐集到該個人的資料。不當的匿名化將會對於資料主體產生直接、間接的負面影響，尤其，當攻擊者（attacker）懷有惡意時[235]。

1. 以殘留的資料指紋再識別化

針對上開案例，如何再識別化 William Weld、Netflix Prize 等資料？Paul Ohm 指出[236]，研究人員通常結合 2 個資料集（每個資料集對於「這資料究竟描述誰？」的問題均提供部分答案），及結合的資料回答或幾乎回答了該「描述誰」的問題。

縱資料管理者已去除其認為可能用以獨特識別出個人之資料欄位，但研究人員仍藉由發現資料中殘留的獨特性而揭開身分。如同遺留於犯罪現場的人類指紋可獨特地識別出某一個人，得以將該人與匿名資料連結在一起；類似地，由一般人所分享資料值之相結合，資料主體亦產生「資料指紋」（data fingerprints）遺留效果。可以比預期更容易之方式而在非個資之上發現資料指紋，乃破壞了匿名現狀。如 Latanya Sweeney 發現，美國人口資料中郵遞區號、生日與性別的驚人獨特性；Arvind Narayanan 與 Vitaly Shmatikov 則揭開，一個人所評分電影資料集的驚人獨特性。據此，對於能接觸適當外部資料之人，所有事務將均是可識別[237]或再識別之個資。

2. 殘留資料致身分揭露而再識別化

當攻擊者能將某一特定資料連結至某一個人，將發生身分揭露（identity disclosure），主要發生於三種情形[238]：

---

235 *Opinion 05/2014 on Anonymisation Techniques* 23.

236 Ohm, *supra* note 120, at 1723.

237 *Id*.

238 GARFINKEL, *supra* note 5, at 12.

（1）不完備的去識別化（insufficient de-identification），指識別資料可能不慎殘留於去識別化資料集，加上不完備的安全管控措施。如前揭 AOL 所釋出關鍵字搜尋的資料檔案，去除某些識別資料但仍保留用戶所輸入的搜尋條件，並輔以隨機分派代碼識別用戶。雖搜尋關鍵字本身（如搜尋財產）未出現識別資料，但代碼比對關鍵字仍能區別出不同的用戶，故記者才能識別出並接觸某些用戶。又如由華盛頓州立醫院釋出去識別化檔案比對導致住院意外的報載，而再識別化[239]。

（2）以連結而再識別化（re-identification by linking），指藉由某些殘留資料與另一可識別資料集中類似特徵（attributes）相連結，而再識別化某些特定檔案。如關鍵字搜尋的檔案雖去除用戶名字，仍留下 IP 地址，而能連結至某一資料庫而找到名字。

（3）假名逆轉（pseudonym reversal），指可識別資料雖假名化，仍可能逆轉該假名化。如研究人員檢視包括 1.73 億筆紐約市計程車路程與匿名化駕照、計程車編號及其他元資料在內的 20GB 資料集之後，能逆轉假名而輕易再識別駕駛人身分[240]，故所採去識別措施其實未能保護隱私。

（三）再識別化之動機

對於去識別化資料，進行再識別化攻擊的各種可能動機、理由，如下[241]：

1.測試去識別化之品質，如受資料控管者委託，研究者進行再識別化攻擊，惟事前應簽署保密協定、採行適當安全維護措施。

2.擁有可進行再識別化或甚可對之加以公開的專業立場，如有些成功的再識別化具有新聞價值或有研究專業上益處，故倘契約或法律未加限制，則可合法進行攻擊。

---

239 LATANYA SWEENEY, MATCHING KNOWN PATIENTS TO HEALTH RECORDS IN WASHINGTON STATE DATA, https://privacytools.seas.harvard.edu/files/privacytools/files/1089-1.pdf (last visited Apr. 14, 2022).

240 Rostow, *supra* note 26, at 689-90; GARFINKEL, *supra* note 5, at 12, 17.

241 GARFINKEL, *supra* note 5, at 10.

3.羞辱或傷害進行去識別化的機構，如進行去識別化的機構常承擔保護個資義務，倘能證明「所採保護措施不夠的」，即可羞辱或傷害該機構，尤其，當去識別化資料曾被公開釋出。

4.直接由再識別化獲利，如行銷業者可能購買去識別化健康資料，以與身分進行配對，而將處方藥折價券寄予被再識別出之個人。再者，惡意競爭對手（如身分竊盜）亦藉出售再識別化檔案而獲利，如英國病歷曾委外印度進行電腦化，卻於黑市銷售而每一檔案索價 4 英鎊，當資料集整體出售則有更大用途（如追蹤藥物濫用）而可索價更高（如駭客曾向維吉尼亞州衛生當局索價 1,000 萬美元）[242]。

5.製造問題，倘敏感資料再識別化而被知悉身分，造成羞辱、勒索或傷害等負面後果。

當前去匿名化漸可行，乃有強烈財務誘因再識別化消費者資料，以供針對性行銷。再識別化乃成有利可圖生意，惟有隱私侵害風險[243]；故有主張[244]，應以契約、立法（本文伍之三部份詳述）管控或禁止再識別化。

（四）再識別化之風險評估

理想上，有效的去識別化、匿名化應建築在對於「個資構成要件」的完整認識之上。進而言之，個資可謂「涉及可使某一現存個人被識別出來的下列資料：（a）某資料、或（b）該資料與資料控管者現有或將來可能擁有之其他資料。[245]」按此定義，倘某資料本身或結合其他資料而未能識別某一個人，則非屬個資。乍看之下，似可輕易判斷某一特定資料是否與某人有關而

---

[242] EL EMAM & ARBUCKLE, *supra* note 25, at 34-35.

[243] Brasher, *supra* note 10, at 226.

[244] EL EMAM & ARBUCKLE, *supra* note 25, at 26.

[245] Data Protection Act 1998, § 1 (Eng.), https://www.legislation.gov.uk/ukpga/1998/29/section/1/enacted (last visited Aug. 17, 2023).

成為個資。然而，在判斷其他資料是否存在、誰可近用、是否用於再識別[246]，卻非易事。

如前揭，再識別化攻擊，常出自同時擁有「去識別化資料集」與「某些額外背景資料」之攻擊者。尤其，當有公開釋出之資料集，則資料攻擊者毋庸經授權即可取用。因此，為確保去識別化之有效性[247]，應審酌「得由去識別化資料與額外資料而知悉關於某個人之識別符號」之再識別化風險。

由於去識別化、匿名化資料常帶有再識別化風險，倘有再識別化的合理可能性，將衍生某些嚴重的隱私風險。而隱私風險的評估，應按個案的情境脈絡而考量下列因素，（1）資料倘被再識別化、揭露而可能被洩漏的活動，因而造成資料主體的負面效應；（2）資料主體、競爭對手及其他利害關係人相關之角色、關係及權力結構；（3）所應適用的隱私相關規範；（4）相關研究的廣泛價值（宗旨、目標）及對於資料主體的益處[248]。

1. 再識別化風險的情境脈絡評估因素

欲評估量化衡量再識別化風險，乃相當複雜的，因再識別化能力應端視原始資料集、去識別化技術、攻擊者技術能力與可利用資源、可供取用之其他可與去識別化資料進行連結之資料[249]。簡言之，再識別化風險乃按資料內容與處理情境脈絡而定[250]。進一步言，再識別化風險的情境脈絡評估（contextual evaluation）應考量因素包括：資料處理計畫目的、資料本身的揭露風險、近用分析資料之人、資料集如何連結、資料集敏感性程度、資料近用環境（包括物理、技術與程序性的近用管制）、分析結果如何揭露、資

246 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 11, 18.

247 GARFINKEL, *supra* note 5, at 10.

248 Tene & Polonetsky, *supra* note 212, at 258; IAN BROWN, JOSS WRIGHT & DAVID ERDOS, ETHICAL PRIVACY GUIDELINES FOR MOBILE CONNECTIVITY MEASUREMENTS 30-31, (Bendert Zevenvergen 2013) https://www.freehaven.net/anonbib/cache/ZevenbergenBrownWrightErdos2013.pdf (last visited May 29, 2023).

249 *Id*.

250 ARBUCKLE & EL EMAM, *supra* note 5, at 52.

料分析結果將分享或公開予何種對象、與加諸於資料使用者之信賴與管制程度。其中，就資料近用的類型而言，（1）資訊公開，由於對資料的散佈並不加以限制，將有較高風險；（2）以法律上具有可執行的方式而限制資料的分享，將有中等或較高風險，例如，與同一研究聯盟分享資料將有中等風險，而與業者或政府分享資料將有較高風險，惟資料接收者倘為來自同一機構的個別研究者則將有較低風險；（3）受控管資料（Managed Data），將有較低風險；（4）互動模式（Interactive methods），即資料集統計資料的散佈，將有較低風險。另外，可以根據風險程度而將攻擊者、競爭對手分為下列三種類型：（1）檢察官風險（Prosecutor risk）（將有較高風險），即意圖再識別化某一特定個人、擁有可以相結合而洩漏關於資料主體資料之輔助資料（auxiliary data）、擁有法律權限而能強制提供被儲存的資料；（2）新聞記者風險（Journalistic risk）（將有中等至高風險），即搜尋資料集中的某一特定目標、擁有可以相結合而洩漏關於資料主體資料之輔助資料；（3）行銷業者風險（Marketer risk）（將視多少個人被再識別出來而有較低至中等風險），競爭對手意圖識別儘可能多的人，越多人被識別則風險越高。當然，風險的評估，亦須考量競爭對手預期所能投入的能力、技術、時間或所能取用的輔助資料而進行再識別化。例如，美國政府情報業務耗費數十億美元而成為極具能力與較高風險的競爭對手；新聞記者與行銷業者的預算皆較低而動機卻有所不同，行銷業者可能比新聞記者更具能力，但新聞記者可能更有決心再識別化某一個人而投入更多努力[251]。

2. 不易評估資料未來相結合的再識別化風險

[251] VICTORIA STATE GOVERNMENT, DE-IDENTIFICATION GUIDELINES 15 (2018), https://www.vic.gov.au/sites/default/files/2019-03/Victorian-Data-Sharing-Act-2017-De-identification-guidelines.pdf (last visited Nov. 30, 2022); BROWN ET AL., *supra* note 248, at 15, 28-30.

就其他可進行連結之資料，除如前揭須判斷資料供誰（某特定人或任何人［如資料公開上網］）取用外，由於不易精準確認何種資料已供取用或未來會釋出，因此，根本難以預測「以此種資料連結而再識別化之風險」為何[252]。

再識別化風險有時會隨著時間拉長而增加，因隨著技術進步與有更多相關情境脈絡資料（contextual information）可供取用（如資料公開），而不太可能在演算法上能事先預測何種背景資料未來可用於再識別化。尤其，在大數據環境中，資料來源多元，去識別化資料仍可能與他處資料相結合，而難以絕對確定「不可能再識別」。然而，仍應採取各種方法，盡力評估、降低再識別化風險[253]；在創造、分享去識別化資料的一開始，即應善盡注意義務並持續性進行整體風險分析[254]，以減緩隱私侵害。

3. 敏感性資料風險較高而應受較嚴格近用限制

資料集常有不同程度的敏感性而有不同的風險，某些類型資料（如健康、生物、財務資料等）對於當事人較具敏感性而較易形成損害，可能較易為攻擊者鎖定而有較高的再識別化風險。倘將所有類型資料均一視同仁而賦予同等敏感性等級，將犯過度簡化之謬誤[255]。因此，對於敏感性資料的大規模利用而應設定較嚴格標準，歐盟 GDPR 第 35 條即強制要求應進行資料保護影響評估（data protection impact assessments）；美國科羅拉多州與維吉尼亞州亦要求對於敏感性資料或其他資料的處理進行評估[256]。其實，隱私影響評估規模須根據潛在隱私風險而定。倘初步評估認為係較低隱私風險，則簡單的隱私影響評估即足夠；但倘被認為係較高隱私風險，如對於敏感性資料或形

---

252 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 18.

253 GARFINKEl, *supra* note 5, at 10.

254 THE INFORMATION COMMISSIONER'S OFFICE , *supra* note 20, at 18; Anne S.Y. Cheung, *Moving Beyond Consent for Citizen Science in Big Data Health and Medical Research*, 16 NW. J. TECH. &INTELL. PROP. 15, 36 (2018).

255 Jonathan Deitch, *Protecting Unprotected Data in Mhealth*, 18 NW. J. TECH. &INTELL. PROP. 107, 119 (2020)；Rubinstein & Hartzog, *supra* note 69, at 741.

256 Colorado Privacy Act, COLO. REV. STAT. § 6-1-1309(2); Consumer Data Protection Act, Va. Code Ann. § 59.1-580(A).

成大量侵害風險，則應需要較為廣泛的系統性隱私影響評估[257]。而亦有主張[258]，應根據資料敏感程度（可分為高度敏感、敏感、輕微敏感、無敏感資料）而採行符合比例原則之安全與隱私保護程度與措施（如去識別化措施、近用與管控程度）。因此，對於由敏感性資料衍生而來或存有被再識別化高風險的匿名化資料近用特別限制，乃適當之舉。尤其，資料的識別或再識別化可能造成逮捕、虐待、死亡、威脅或長期歧視之極端情形時，應採行不傷害原則（do no harm principle），即不論資料的效用性為何，資料的去識別化程度應達到現實操作上（realistically）不可能被再識別化；倘操作上不具有可行性，則資料完全不應被蒐集[259]。

四、再識別化之負面效應

對於個資去識別化、匿名化後所形成匿名狀態的維護，乃保障隱私權與表意自由的關鍵[260]，以確保交易或政治自由[261]。惟由再識別化實例可知，即使欠缺對於某個人身分的認識，藉由大數據運作仍可建構與其相關的詳細描繪、剖析，而阻礙徹底的匿名化[262]，將侵害個資自主控制權與隱私權。如同

---

257 INFORMATION AND PRIVACY COMMISSION NEW SOUTH WALES, A GUIDE TO PRIVACY IMPACT ASSESSMENTS 5 (May 2020), https://www.ipc.nsw.gov.au/sites/default/files/2021-3/Guide_to_Privacy_Impact_Assessments_May_2020.pdf (last visited Oct. 14, 2022); Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan & Urs Gasser, *Towards A Modern Approach to Privacy-Aware Government Data Releases*, 30 BERKELEY TECH. L.J. 1967, 2063 (2015).

258 Andrew B. Serwin, *Privacy 3.0: the Principle of Proportionality*, 42 U. MICH. J. L. REFORM 869, 900 (2009); VICTORIA STATE GOVERNMENT, *supra* note 251, at 19.

259 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 37; BROWN ET AL., *supra* note 248, at 31.

260 ARTICLE 19, RIGHT TO ONLINE ANONYMITY 1 (June, 2015), https://www.article19.org/data/files/medialibrary/38006/Anonymity_and_encryption_report_A5_final-web.pdf (last visited Apr. 14, 2022).

261 A. Michael Froomkin, *Flood Control on the Information Ocean: Living with Anonymity, Digital Cash, and Distributed Databases*, 15 J.L. & COM. 395, 408-09 (1996).

262 Pan, *supra* note 222, at 256.

Paul Ohm 示警，再識別化科學打亂隱私規範架構，弱化匿名化本可帶來的信任感；藉由顯已去識別化片段資訊，再與其他資訊交叉比對，所形成累積效應（incremental effect）將一點一滴吞噬個人隱私，直到其輪廓、描繪完全被揭穿、曝露為止[263]，造成隱私等人格與經濟損害及表意自由的寒蟬效應。

（一）隱私等人格與經濟損害

再識別出的資料，倘包括敏感資料而讓名字、位址等重要資訊為人所知，將造成隱私等人格損害。例如，傳染病資料之蒐集、善用而增進公衛，惟人們對於傳染性性病常諱疾忌醫，更擔憂被識別出罹患 AIDS[264]。倘資料連結至被研究者名字或地址，並指出住處受污染，可能影響房價致經濟損失[265]。

（二）表意自由之寒蟬效應

匿名性是否為一件好事，並無共識。有主張，藉由匿名化而逃避不法行為之偵測而得從事有害行為（如毀謗、智財侵害），故應禁止某些形式的匿名性[266]。惟能夠在線上匿名化，應是諸多能構成網路特性（如能坦率評論、對於威權政府之異議、吹哨而揭黑幕）的一切基石。美國聯邦最高法院[267]即肯認，拒絕揭露社員名單、匿名撰寫傳單，均為憲法保障之權。曝露所屬團體與作者身分，將對於結社與表意自由形成寒蟬效應。惟縱去識別化、匿名化資料，在大數據環境中，得萃取或推衍出線上使用者身分，將降低毫無顧

---

263 Paul Ohm, *Don't Build a Database of Ruin*, HARVARD BUSINESS REVIEW (Aug. 23, 2012), https://hbr.org/2012/08/dont-build-a-database-of-ruin (last visited Apr. 14, 2022).

264 Froomkin, *supra* note 261, at 408-09.

265 Latanya Sweeney, Ji Su Yoo, Laura Perovich, Katherine E. Boronow, Phil Brown & Julia Green Brody, *Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study*, TECHNOLOGY SCIENCE (Aug. 27, 2017), https://techscience.org/a/2017082801/ (last visited Apr. 14, 2022).

266 Froomkin, *supra* note 261, at 402-07.

267 NAACP v. Alabama, 357 U.S. 449, 462-63 (1958); Talley v. California, 362 U.S. 60, 64-65 (1960).

忌表達意見、參與互動或進行交易之意願[268]。為避免新穎監控科技削弱言論自由的保護，法律須因應調整[269]。乃有主張，為確保免於被識別、追蹤、剖析[270]的匿名權利，除對於不當行為（如毀謗、詐欺、霸凌、網路騷擾［cyber-stalking］、與兒童不當互動等侵權）之執法外，否則，利用資料而重新再識別本欲匿名之當事人，即可能屬於一種侵害[271]。

五、小結

一些再識別化事件，使得去識別化的有效性漸受質疑。但有反駁，被成功再識別出來的比例微小，去識別化仍為有效機制。

當資料單獨存在且未能連結至某一個人，去識別化是可行的。不過，大數據環境中，去識別化漸失效，因去識別化資料仍可能與其他來源資料相結合，而難以絕對確定不會被再識別出來，進而曝露於隱私侵害風險，以及形成人格與經濟損害及表意自由的寒蟬效應。

再識別化已成一門有利可圖生意，乃有主張，應管控資料接收者再識別化動機，並應採取各種方法而降低風險。

肆、對於去識別化有效性之爭論

對於去識別化、匿名化做為隱私保護之有效性、可能性，法律與隱私學者、專家見解分歧而展開涇渭分明的學術辯論，並導出不同的規範政策取向[272]。類似地，不同法院之間見解亦相當分歧。

---

[268] Pan, *supra* note 222, at 256.

[269] Julie E. Cohen, *A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace*, 28 CONN. L. REV. 981, 1012 (1996).

[270] Rubinstein, *supra* note 195, at 78.

[271] PCAST, *supra* note 9, at 8.

[272] Brasher, *supra* note 10, at 231-32.

一、對於去識別化、匿名化有效性之學界見解分歧

（一）對於去識別化、匿名化有效性之否定見解

1. Paul Ohm「匿名化不可避免將失靈」之觀點

Paul Ohm 在 2010 年深具影響力文章「隱私的破碎承諾：匿名化意想不到失靈之因應」，率先將電腦次領域「再識別化科學」引入法學，在法律與決策圈引起激辯，究竟應如何因應「電腦科學關於再識別化之研究成果」？該文揭穿「資料乃有用或完美匿名，二者不能並存」之緊張關係[273]。

對於匿名化之效用，Paul Ohm 文章採取批判觀點，將匿名化（其稱為「釋出後不理的匿名化」[release-and-forget anonymization][274]）視為一種愚人黃金（fool's gold），因當前匿名化方法已不太能發揮保護隱私功效。電腦科學進步漸使再識別化技術普及化，使得攻擊與再識別出理論上本應屬匿名化資料變得可行，亦使得本意圖以匿名化保護隱私[275]之功效備受質疑[276]。因此，Paul Ohm 認為，匿名化無疑是一種失靈，故應放棄「將匿名化視為隱私保護唯一、排他手段」[277]，至少應放棄視「僅藉由移除可識別個人資料（PII）即可保護隱私」為理所當然的觀點，匿名化不應再視為可提供隱私之有意義保證或萬靈丹[278]。

Paul Ohm 再指出[279]，對於匿名化技術確保隱私的能力，現行隱私法規顯然過於樂觀。技術、執法人士皆誤信「資料輕微變動，即可有效保護隱私」。電腦科技理論已建立「任何即使用途短暫的資料，亦不可能完美匿名化；效

---

273 Ohm, *supra* note 120, at 1703-04, 1742-43.

274 所稱「釋出後不理的匿名化」，即資料管理者進行匿名化處理後，釋出檔案予第三人或僅予組織內部，即忘卻此事，未曾試圖追蹤檔案後續狀況。*Id*. at 1711-12.

275 Peppet, *supra* note 224, at 129.

276 Donna M. Gitter, *Informed Consent and Privacy of Non-Identified Bio-Specimens and Estimated Data: Lessons from Iceland and the United States in an Era of Computational Genomics*, 38 CARDOZO L. REV. 1251, 1280 (2017).

277 Rubinstein & Hartzog, *supra* note 69, at 723.

278 Brasher, *supra* note 10, at 231-32; Gitter, *supra* note 276, at 1280.

279 Ohm, *supra* note 120, at 1704, 1706-07, 1751, 1755.

用的微小增進，將導致隱私巨大損失」；隱私與效用二者根本不相容[280]。為了保有用處，資料須不完美匿名化。完美的隱私固可藉由完全不揭露任何事務而達成，但將不利於效用；完美的效用固可完全按接收者所需程度而揭露資料，但不利於隱私。其實，不論如何進行資料的匿名化，對手倘擁有適當的外部資料，仍可利用資料的殘餘效用（residual utility）而進一步揭露其他資訊。社會常存在某些蛛絲馬跡的外部資料而可與匿名化資料相結合，藉以揭露私人資訊。因此，對於保有用處的資料，完美的匿名化是不可能的[281]。

Paul Ohm 乃指出[282]，再識別化的容易化，顛覆奠基匿名化所建構之隱私保護承諾與期待；「能夠有效匿名化」之誤認已瀰漫整個法律體系，造成制度浩劫，實應採行適當因應之道。

近來隱私的辯論，Paul Ohm 上開觀點常被引用而深具影響力，用以支持「效用與匿名化根本不相容」之論點[283]。

2. 其他學者亦質疑去識別化、匿名化功效之見解

類似於 Paul Ohm，有些學者亦質疑去識別化、匿名化之功效。其中，Latanya Sweeney 主張[284]，縱利用去識別化技術，亦不能將隱私侵害可能性從公開揭露的資料集中完全消除，因包含相關資料的其他資料集不可避免亦將被揭露，兩相連結，可再識別出第一個資料集中的特定個人。

Arvind Narayanan 與 Vitaly Shmatikov 指出，去識別化（匿名化）功效越來越差，乃因植基於可識別（identifying）、非可識別（non-identifying）特徵之錯誤分類；該分類，針對原始狀態下的攻擊（original attack）是有用的，但隨著可公開取得關於個人的資料數量、種類快速增加而漸失意義。因

---

280 Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 U. COLO. L. REV. 1117, 1126-29 (2013).

281 Ohm, *supra* note 120, at 1752.

282 *Id*. at 1707.

283 Wu, *supra* note 280, at 1127.

284 Latanya Sweeney, *k-Anonymity: A Model for Protecting Privacy*, 10 INT'L J. UNCERTAINTY, FUZZINESS AND KNOWLEDGE-BASED SYS. 557, 558-59 (2002); Rubinstein & Hartzog, *supra* note 69, at 704-28.

此，去識別化僅能提供一種較脆弱形式的隱私保護，或可防止內部人窺探（“peeping” by insiders）及讓誠實人繼續維持誠實而已[285]。

Omer Tene 與 Jules Polonetsky 指出，去識別化資料僅為暫定狀態，而非穩固分類；縱已去除可供識別的資訊，仍可由大數據演算法推論出身分[286]，阻礙徹底去識別化（匿名化）[287]。AI 尤擅長由看似不相干資料萃取關聯性而進行再識別化[288]。John P. Dever 與 Captain James A. Dever 乃指出「資料今日縱經充分去識別化，明日仍可能被再識別化」之隱憂[289]。

此外，其他亦有質疑去識別化、匿名化功效之見解，如（1）N. Nina Zivanovic 指出，去識別化資料分開個別檢視是安全的，但當與各種公開資料一併檢視，則可再識別化[290]。（2）Amy Westergren 指出，可識別之資訊可能意外繼續附著於釋出的去識別化資料，個人身分可能藉由再識別技術加以揭露[291]。（3）John Verdi 指出，病歷資料包含諸多資訊（如地點、性別、醫療狀況、年齡），當結合在一起，加上強大電腦運算能力，可再識別病人[292]。（4）Scott R. Peppet 指出，物聯網提供多元化資料，欲以去識別化保護隱私越來越難[293]。

3. 政策制定者亦質疑去識別化、匿名化功效之見解

---

285 Narayanan & Shmatikov, *supra* note 1, at 24, 25-26.

286 Tene & Polonetsky, *supra* note 212, at 257.

287 Pan, *supra* note 222, at 256.

288 Cunningham, *supra* note 204, at 537; Manheim & Kaplan, *supra* note 198, at 128.

289 John P. Dever & Captain James A. Dever, *A Democracy of Users*, 6 J.L. & CYBER WARFARE 8, 27 (2017).

290 Zivanovic, *supra* note 53, at 191.

291 Amy Westergren, *The Data Liberation Movement: Regulation of Clinical Trial Data Sharing in the European Union and the United States*, 38 HOUS. J. INT'L L. 887, 898 (2016).

292 John Verdi, *Transcript: Sorrell v. Ims Health-Any Impact on Patient Privacy?*, 36 VT. L. REV. 829, 830 (2012).

293 Peppet, *supra* note 224, at 130-32.

Paul Ohm 等學者不斷主張「資料未能可靠地被去識別化」，政策制定者開始相信相關論述並抱持類似見解[294]。例如，主管 HIPAA 的美國 OCR 提出，去識別化從非完美的，去識別化常僅是暫定而非永久的狀態。縱按 HIPAA 隱私規則（Privacy Rule）經適當去識別化資料，仍可能帶有私人資訊，足以形成某些再識別化風險，亦即未來足以滋長、擴大之風險；換言之，當資料集被釋出而與其他資料集結合、比對，將解開一個人身分的關鍵而再識別化[295]。前揭白宮大數據與隱私相關報告[296]亦有類似質疑[297]。

FTC 提出警告[298]，去識別化資料可能會被再識別化，使得可識別個人資料（PII）與非可識別個人資料（non- PII）之分野，漸失重要性；乃有指出「只要有充分時間與資源，任何資料均可能連結至某一個人」。

（二）對於去識別化、匿名化有效性之肯定見解

對於去識別化、匿名化有效性之爭論，Jane (Yakowitz) Bambauer （原名為 Jane Yakowitz）則為辯論相反一邊代表人物。Jane Bambauer[299]雖亦承認去識別化（匿名化）有侷限性，但認為 Paul Ohm 等專家誇大再識別化風險之危害性，卻低估資料對外釋出的價值[300]；Jane Bambauer 主張，資料流通的益處實大於去匿名化的風險性[301]。再者，對於知名的再識別化攻擊案例之精確性與正當性，漸有質疑聲浪；Jane Bambauer 與 Daniel Barth-Jones 即主張，知名的再識別化攻擊三重奏（Weld、AOL、Netflix）扭曲了政策辯論，因其不具代表性且已受大眾媒體扭曲。如同 Paul Ohm，Jane Bambauer 與

---

294 Ann Cavoukian & Daniel Castro, *Big Data and Innovation, Setting the Record Straight: De-identification Does Work* 2 (2014), http://www2.itif.org/2014-big-data-deidentification.pdf (last visited Apr. 14, 2022).

295 NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS, *supra* note 37, at 5, 8.

296 EXECUTIVE OFFICE OF THE PRESIDENT, *supra* note 200.

297 Cavoukian & Castro, *supra* note 294, at 2.

298 FEDERAL TRADE COMMISSION, *supra* note 75, at 19-20.

299 Yakowitz, *supra* note 7, at 4-5, 42.

300 Wu, *supra* note 280, at 1140.

301 Brasher, *supra* note 10, at 232.

Daniel Barth-Jones 雖亦肯定該等攻擊而得以彰顯去識別化技術之不足，惟應以其他方式加以詮釋[302]。

1. 再識別化風險實屬微小

Jane Bambauer 認為，再識別化風險主要屬理論、假設層次，言過其實，現行一些去匿名化、再識別化例子乃異常值（outliers）、特例；其實，再識別化之隱私風險，實低於許多日常視為理所當然的風險承擔（如丟棄垃圾可能造成隱私外洩）[303]。其乃主張，匿名化資料流通應更容易而非更困難[304]，以供研究之用。

去識別化（匿名化）的其他捍衛者亦認為，儘管理論上及已經證明有再識別化攻擊能力，惟對大多數資料集，再識別的可能性仍屬微小[305]。Khaled El Emam 與 Luk Arbuckle 即認為，倘以周延方式進行健康資料的去識別化，再識別化的風險可能非常微小；知名的再識別化案例，其實非針對經適當去識別化資料而為之。故倘有認為「縱非全部，大部分仍可能再識別化」，實為一種迷思[306]。Daniel A. Moros 與 Rosamond Rhodes 亦主張，並無證據或顯著理由可證明，當前對於去識別化的保護有所不足[307]。

(1) 再識別化需各種資源而非易事

某些學者則認為，由於受到要達能夠再識別化所需努力的箝制，資料再識別化風險其實小於一般的認知[308]。Daniel Barth-Jones 即主張，即使借助電腦的可觀協助，在渾沌的資訊之海，人類仍無足夠時間與精力而能進行追蹤、釐清與查證，以在茫茫人海中明確再識別出某一個人；至少倘有適當的去識

---

302 Rubinstein & Hartzog, *supra* note 69, at 723-24.

303 Yakowitz, *supra* note 7, at 39-40.

304 Rubinstein & Hartzog, *supra* note 69, at 724.

305 *Id.* at 705.

306 EL EMAM & ARBUCKLE, *supra* note 25, at 9.

307 Barbara J. Evans, *Why the Common Rule Is Hard to Amend*, 10 IND. HEALTH L. REV. 365, 414 (2013).

308 Gitter, *supra* note 276, at 1280.

別化方式，將使得再識別化的成功機會微小。Daniel Barth-Jones 與 Jane Bambauer[309]更指出，試圖去匿名化的攻擊未能大規模可行，因一般群體（general population）屬性、特徵的確認乃非同小可，常超乎資料入侵者可承受範圍。因每次攻擊須針對去識別化資料庫與群體量身打造，以求能如同重現資料蒐集當時狀況，故只有針對非比尋常情形下的小規模群體，攻擊才行得通[310]。

加拿大資訊保護官 Ann Cavoukian 與資深分析師 Daniel Castro[311]亦指出，進行再識別化，常需有其他替代來源的資料可與現有具有可識別特徵的資料進行比對，才能成功再識別化；惟資料取得未必容易。Daniel C. Barth-Jones 亦認為[312]，難以取得背景、輔助資料而未能使資料主體可被辨識或具有獨特性，故未能再識別化。

Ann Cavoukian 與 Daniel Castro[313]另指出，再識別化需具高度訓練、技巧的專業知識，故再識別化的真實風險可能遠低於預期。如同 *Southern Illinoisan v. Department of Public Health*[314]案法官指出，專家能夠操縱資料而查明身份，非必然意謂「其他人也有此種能力」之威脅確實存在著，或即使威脅確實存在著，亦非必然意謂「資料的釋出，即合理地會導致特定個人身

---

309 JANE YAKOWITZ & DANIEL BARTH-JONES, THE ILLUSORY PRIVACY PROBLEM IN SORRELL v. IMS HEALTH 1, 7 (May 2011), https://techpolicyinstitute.org/wp-content/uploads/2011/05/the-illusory-privacy-problem-i-2007545.pdf (last visited Apr. 14, 2022).

310 Gitter, *supra* note 276, at 1280-81.

311 Cavoukian & Castro, *supra* note 294, at 2, 5.

312 Daniel C. Barth-Jones, *Public Policy Considerations for Recent Re-Identification Demonstration Attacks on Genomic Data Sets: Part 1 (Re-Identification Symposium)*, BILL HEALTH HARV. L. BLOG (May 29, 2013), https://blog.petrieflom.law.harvard.edu/2013/05/29/public-policy-considerations-for-recent-re-identification-demonstration-attacks-on-genomic-data-sets-part-1-re-identification-symposium/ (last visited Aug.. 17, 2023).

313 Cavoukian & Castro, *supra* note 294, at 2, 5.

314 Southern Illinoisan v. Department of Public Health, 349 Ill. App. 3d 431, 812 N.E.2d 27 (Ill. App. Dist.5 2004), https://casetext.com/case/the-southern-illinoisan-v-deptof-public-health (last visited Apr. 14, 2023).

份之識別」。Felix T. Wu 乃認為[315]，Paul Ohm 誤解電腦科學理論的某些重要觀點而主張「隱私與效用根本不具相容性」。

(2) 經適當去識別化後，再識別化比例可能非常低

Donna M. Gitter 認為，面對資料效用與再識別化風險二者價值衝突，現行去識別化標準常錯誤地與隱私站在同一邊。例如，HIPAA 安全港模式嚴格要求去除 18 種識別符號而進行去識別化；惟某些實證研究，對於 HIPAA 再識別化真實風險進行評估，而證明過於誇大其辭。其中，有研究，15,000 個病人檔案按 HIPPA 現行嚴格規定而進行去識別化後，嘗試與其他可取得外部資料進行比對，卻僅有 2 個人被精確再識別出來，再識別率僅有 0.013%[316]；另一項研究，按 HIPPA 去識別化的 128,000 人資料中，亦僅有 2 個人被再識別出來[317]；或按另一專家評估，依安全港標準進行去識別化的資料集中之個人，僅有 0.04%具有獨特的可識別性[318]。或其他研究顯示，再識別化風險將因州別不同而有差異，懷俄明州為 0.25%、加州則為 0.01%，相差至少 25 倍[319]。另外，在 Netflix 的 480,189 個用戶中，僅能再識別化其中 2 個用戶（0.0004%）[320]。Daniel C. Barth-Jones 亦指出，Paul Ohm 前揭文章所提之攻擊案例（Weld, AOL, Netflix），在 125 萬人中，僅有 4 人被識別出；因此，Paul Ohm 所謂「科學家已證明，可輕易進行再識別化」主張，令人存疑[321]。

---

315 Wu, *supra* note 280, at 1129.

316 Gitter, *supra* note 276, at 1281-82; Rubinstein & Hartzog, *supra* note 69, at 711.

317 DANIEL C. BARTH-JONES, WHY A SYSTEMS-SCIENCE PERSPECTIVE IS NEEDED TO BETTER INFORM DATA PRIVACY DE-IDENTIFICATION PUBLIC POLICY 9, https://fpf.org/wp-content/uploads/2016/11/DB-J-Brussels-Privacy-Symposium-Systems-Science-Perspective-for-Data-Privacy-11_10_16-Final.pdf (last visited Apr. 14, 2022).

318 Cavoukian & Castro, *supra* note 294, at 5.

319 Benitez & Malin, *supra* note 168, at 169.

320 Gitter, *supra* note 276, at 1281-82; Rubinstein & Hartzog, *supra* note 69, at 715.

321 BARTH-JONES, *supra* note 317, at 9, 14.

可知，經適當去識別化資料集，縱以極端努力，被再識別化風險卻可能仍非常低；因此，其關鍵，應非去識別化是否有效（其實，確實有效），而是能否有效善用去識別化（但其實，並非總能有效善用）[322]。

2. 過於捍衛去識別化資料而限制釋出，將導致資料公地之悲劇而不利公益

一些務實的電腦科學、統計與流行病學家認為[323]，有用資料的分享有助於公益，乃紛紛倡議某些去識別化的方法與因應之道。

Jane Bambauer 認為，Paul Ohm 等論者忽略資料公地、公共財（data commons）之價值，Jane Bambauer 將資料公地形容為「人們所支付予公共資料保留庫（public information reserves）之稅捐」[324]，可造福各領域。其認為，捍衛匿名化資料而限制或終結對外釋出，將導致新型態的資料公地悲劇，使社會整體受害[325]；換言之，當資料主體藉由移除其個資而使公地、公共資源枯竭時，個人雖享有可避免被再識別化風險之完整利益，但其決策的邊際危害將產生外部效果[326]而由整個群體共同承擔[327]。因此，Jane Bambauer 主

---

322 Cavoukian & Castro, *supra* note 294, at 8.

323 Rubinstein & Hartzog, *supra* note 69, at 715.

324 Yakowitz, *supra* note 7, at 2-3, 66.

325 Rubinstein & Hartzog, *supra* note 69, at 724.

326 對此，有主張，許多政治經濟情境能以賽局中囚徒困境加以分析。針對參與者超過兩方的公共財賽局（public goods game），如燈塔、道路、學校等公共財能造福社群內所有人，若由別人出錢建設而自己免費享用，利益自然更大，燈塔蓋好沒辦法只照亮特定人。因此，透過強制誓約可避免囚徒困境中相互背叛的惡果，具有強制力法律與契約也能藉由懲罰而讓大眾達成公共財賽局內的互利。而社會契約體現公正性的道德邏輯，能抑制惡的誘惑，避免有人成為傻瓜或互相背叛形成悲劇。Steven Pinker（著），陳岳辰（譯）（2022），《理性：人類最有效的認知工具，讓我們做出更好的選擇，採取更正確的行動》，頁303-305，商周。按此主張，得立法強制參與資料公地建構，避免搭便車行為。

327 類似地，國內亦有因立法強制全民納保所全面蒐集建構的全民健康保險研究資料庫運用，而衍生個人得否請求移除其健康保險資料而不參與資料公地建構之爭議。例如，一方面，針對「刪除上訴人等8人資料之資料樣本資料，影響輕微」之主張，最高行政法院106年度判字第54號判決認為：「但如果容許少數人退去，基於執法平等性之要求，多數人也可比照辦理，如此可能引發退出風潮，形成『破

窗效應』，造成資料蒐集投入成本之虛耗。」此見解較接近上開公共財賽局得強制參與資料公地建構之主張。

惟另一方面，基於個資自主控制權，不乏主張「當事人應有選擇退出之權」（其實，歐盟GDPR第7(3)條即規定「資料主體有權得隨時撤回其同意」，以落實個資自主控制權），因而質疑強制個人參與資料公地建構而供學術研究之正當性、必要性。如⑴學者劉靜怡指出，除非學術研究具有極高之公共利益重要性甚至緊急性或迫切性，且顯然超過被研究對象個人依其自主性判斷是否認同該等研究之價值的權益，否則，不應強制被研究對象毫無選擇地提供個人資料供研究。參憲法法庭111年憲判字第13號判決劉靜怡教授意見書，頁13。⑵學者李寧修指出，學術研究成果可嘉惠不特定多數人，仍不應以此作為限制當事人個人資料自主控制權之當然、唯一依據，而課以當事人必須容忍之義務。參李寧修（2020），〈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〉，《臺大法學論叢》，49卷1期，頁8、45。⑶學者張陳弘指出，在英國，須監督嚴重傳染疾病或公共衛生嚴重風險的例外情況，才可拒絕當事人行使事後退出權。參張陳弘（2018），〈國家建置全民健康保險資料庫之資訊隱私保護爭議：評最高行政法院106年度判字第54號判決〉，《中原財經法學》，40期，頁247-248。⑷學者劉定基指出，個資法系爭規定，完全剝奪資料當事人自主決定資料是否供統計、學術研究目的外利用，未針對研究之公益性，訂定對抗當事人拒絕權（退出權）之標準，顯非達成重大公益之侵害較小手段，違反比例原則而違憲。參憲法法庭111年憲判字第13號判決劉定基教授意見書，頁20。

對於上開爭議，憲法法庭111年憲判字第13號判決之法律上意見第48段雖肯認「寓有透過統計或學術研究累積科學知識技術等公共財」，但判決主文第四項認定「欠缺得請求停止利用之規定，乃違憲」，而判決：「衛生福利部中央健康保險署就個人健康保險資料之提供公務機關或學術研究機構於原始蒐集目的外利用，由相關法制整體觀察，欠缺當事人得請求停止利用之相關規定；於此範圍內，違反憲法第22條保障人民資訊隱私權之意旨。相關機關應自本判決宣示之日起3年內制定或修正相關法律，明定請求停止及例外不許停止之主體、事由、程序、效果等事項。逾期未制定或修正相關法律者，當事人得請求停止上開目的外利用。」另判決之法律上意見第32段指出「……，資訊隱私權保障當事人原則上就其個資，於受利用之前，有同意利用與否之事前控制權，以及受利用中、後之事後控制權。除當事人就獲其同意或符合特定要件而允許未獲當事人同意而經蒐集、處理及利用之個資，仍具事後控制權外，事後控制權之內涵並應包括請求刪除、停止利用或限制利用個資之權利。」

循此，本號判決黃昭元大法官部分不同意見書第27、30段亦指出：「……，考量人民對於國家之強制蒐集健保資料並建置健保資料庫予以利用，早已無從行使其事前同意權，在制度上自有必要賦予個別人民行使事後之退出權，以此平衡人民事前同意權之喪失，這可說是人民對其個人健保資料自主控制權的最後一道防衛。就權利衡平的抽象審查而言，有個人退出權之保障，才足以維持集體性公益與個

張，正確方向應將匿名化資料定位為可自由分享；過於抗拒資料釋出，將阻礙醫學、公衛與社會科學等重要研究，卻僅能帶來隱私保護些微利益。其實，科技仍可減緩資料效用與揭露風險間的緊張關係，匿名化仍是完美的折衷因應之道。因此，相對於質疑匿名化功效的偏執聲浪，法學界的理性聲音相反地應該喚醒參與資料公地建構的公民義務[328]。

Jane Bambauer 認為，去匿名化、再識別化例子實為特例，簡單的技術即足供保護，而免於被再識別化之真正風險[329]。其批評，當前管制資料釋出之立法，係對於資料利用的掣肘[330]；故其建議，法律應鼓勵而非打擊資料的釋出，可供公共研究的資料分享應更容易，以及應建立「得以用相對明確技術而進行匿名化」之資料釋出安全港[331]。

Jules Polonetsky、Omer Tene 與 Kelsey Finch 亦指出，縱某些可間接識別符號（indirect identifiers）（如日期、地理位置、交易代碼）常用於再識別化攻擊，但得用以揭穿去識別化之資料點，亦常肩負著供重要研究的社會福祉，故不應過度強調再識別化風險[332]；否則，倘資料看起來不涉及個人，但因藉由某些可能方式（不論多麼地不可能加以運用）而再識別、再連結至某

別性私益間之衡平，而不致淪於公益永遠大於私益的空洞衡平想像。」而對於用以反對賦予個人退出權之抽樣偏差、破窗效應主張，黃昭元大法官則反駁：「研究者本就不可能窮盡蒐集擬研究之事物，以追求絕對正確之知識或真理，通常多僅能在特定的研究限制下追求相對正確之知識或真理。以我國而言，有關機關對外提供的資料檔本來就有『兩百萬人抽樣檔』，而非必然是『全人口檔』。相較於臺灣全人口而言，『兩百萬人抽樣檔』本就有其偏差，即使有部分人口選擇退出，究竟會因此增加多大的抽樣誤差，是否因此會造成不可彌補的誤差風險，有關機關對比並未提出具有說服力的資料或證明。何況退出權並非絕對，在制度設計上亦得有不許退出的例外情形。本判決主文第四項亦容許相關機關以法律明定請求停止及例外不許停止之主體、事由等事項，而已留給相關機關相當的制度形成空間。」再者，參考有實施退出權之外國實證經驗（如英國），行使退出權之人數占比約僅略高於5%。

328 Yakowitz, *supra* note 7, at 66-67; Wu, *supra* note 280, at 1122.

329 Yakowitz, *supra* note 7, at 1.

330 Brasher, *supra* note 10, at 232.

331 Wu, *supra* note 280, at 1140.

332 Polonetsky et al., *supra* note 8, at 619.

一個人仍有微乎其微的可能性，而仍應受隱私法規拘束，則資料運用之益處可能受嚴重箝制[333]。

據此，在隱私侵害與促成科學、醫療照護進步福祉之間，去識別化政策須妥為平衡因應[334]。用以消除不確定隱私風險的去識別化規範倘過於嚴苛，可能癱瘓有價值的資料運用，其回報卻是微小隱私利益[335]。

（三）小結

綜上，學界對於去識別化（匿名化）有效性之見解分歧，（1）否定說認為，去識別化資料仍然保留某些得藉以識別之風險，即使風險很小，但仍非等於零；去識別化資料仍可能連結回去資料所對應之當事人。（2）肯定說認為，去識別化雖具有侷限性，縱經極端努力，被再識別化風險可能仍非常低。Paul Ohm 等專家誇大了再識別化風險之危害性而低估資料對外釋出的價值。去識別化仍是可減緩資料效用與揭露風險之間衝突的妥協、折衷之道。

二、美國法院對於再識別化風險之見解分歧

對於個資（如個人健康、醫療資料）之再識別化風險，某些美國法院已依 HIPAA 隱私規則而進行審理，但再識別化之風險評估，法院見解卻相當分歧。

（一）法院認「縱去除 HIPAA 識別符號，仍有再識別化風險與隱私侵害」

在某些案件，法院承認再識別化風險與隱私侵害。*Northwestern Memorial Hospital v. Ashcroft*[336]案法院認為，即使去除 HIPAA 識別符號，病

---

333 Tene & Polonetsky, *supra* note 212, at 258.

334 Gitter, *supra* note 276, at 1282.

335 Polonetsky et al., *supra* note 8, at 619.

336 Nw. Mem'l Hosp. v. Ashcroft, 362 F.3d 923, 929 (7th Cir. 2004).

人的熟人或技巧熟練的 Google 用戶仍可由所揭露的資料而再識別出病人，侵害隱私[337]，因此可能受到威脅、羞辱與公開辱罵[338]。類似地，*Havemann v. Astrue*[339]案法院認為，資料的進一步揭露將導致再識別化與侵害隱私[340]。

Parkson v. Cent. DuPage Hosp.[341]案法院認為，從醫療檔案去除病人名字仍不能保證機密性，因檔案包括病人醫療紀錄，累積的資訊使得識別可能性變得非常高[342]。

（二）法院認「倘去除 HIPAA 識別符號的資料，則屬於充分的去識別化」，而駁回「再識別化風險」之主張

相反地，其他許多法院裁判則駁回「再識別化風險」之主張，因法院認定「倘去除 HIPAA 識別符號的資料，則屬於充分的去識別化」。在 *Baser v. Department of Veterans Affairs*[343]案，法院駁回退伍軍人事務部拒絕依資訊自由法（Freedom of Information Act，下稱 FOIA）提供資料之簡易裁判聲請，即使該部提供專家意見而用以分析「當上開所提供資料與其他可公開或商業上取得資料庫互相連結之再識別化風險」。法院仍主張，只要利用去識別化的兩種方法（專家認定或去除 18 種 HIPAA 識別符號）之一，並未要求應同時利用兩種方法，資料即已完成「去識別化」[344]。

在 *Steinberg v. CVS Caremark Corporation*[345]案，法院駁回原告「依 HIPAA 標準而去識別化的資料，仍可能被再識別化」的主張；因原告僅提供某一學

---

337 Sejin Ahn, *Whose Genome Is It Anyway?: Re-Identification and Privacy Protectionin Public and Participatory Genomics*, 52 SAN DIEGO L. REV. 751, 774-775 (2015).

338 Klocke, *supra* note 166, at 536.

339 Havemann v. Astrue, No. ELH-10-1498, 2012 WL 4378143, at *7-9 (D. Md. Sept. 24, 2012).

340 Ahn, *supra* note 337, at 775.

341 Parkson v. Cent. DuPage Hosp., 435 N.E. 2d 140, 144 (Ill. App. Ct. 1982).

342 Klocke, *supra* note 166, at 536.

343 Baser v. Dep't of Veterans Affairs, No. 13-CV-12591, 2014 WL 4897290, *4-5, *7 (E.D. Mich. Sept. 30, 2014).

344 Ahn, *supra* note 337, at 774.

345 Steinberg v. CVS Caremark Corp., 899 F. Supp. 2d at 339 (E.D. Pa. 2012).

術期刊而解釋再識別化之一般性風險，而非提供某一專家針對本案系爭資料之真實分析[346]。類似地，在 *IMS Health Corp. v. Rowe*[347]案，緬因州司法部長欲證明「居住鄉村的個人被再識別化可能性，相對容易」；惟法院駁回此對病人隱私將有潛在衝擊之主張，因其所強調僅為理論上威脅[348]。

另就其他一般資料，*In Department of Air Force v. Rose*[349]案法院亦認為，刪除涉及個人與其他可識別資料，即足以保護隱私，則按 FOIA，資料可釋出[350]。

（三）小結

綜上，對於個資再識別化風險，法院見解相當分歧。某些案件，法院承認「即使將 HIPAA 識別符號去除而仍有再識別化風險與隱私侵害」。相反地，其他案件，法院則認定「倘去除 HIPAA 識別符號的資料，則屬於充分的去識別化」。

伍、去識別化與再識別化衝突的因應之道

針對上開個資去識別化與再識別化的衝突，可採取下列因應措施，以兼顧資料效用與隱私保護間的平衡、並降低再識別化風險。

---

346 Ahn, *supra* note 337, at 775.

347 IMS Health Corp. v. Rowe, No. cv-07-127-B-W, 2007 WL 4480639, at *13 n.28. (D. Me. Dec. 21, 2007).

348 Klocke, *supra* note 166, at 536.

349 Department of Air Force v. Rose, 425 U.S. 352, 381 (1976).

350 Charkow, *supra* note 58, at 220.

一、資料釋出不免須忍受被再識別化風險，立法亦多採合理可能去識別化的類似標準

對於去識別化資料被再識別化風險，以及對於去識別化、匿名化有效性，由前揭學者與法院分歧見解可知，去識別化、匿名化之個資仍可能與散落他處資料相結合而再識別化，其風險程度縱使微小，仍難以絕對確定「一個人不可能由某一特定資料集而被識別出來」。因此，關鍵之處，應非在於完全排除再識別化風險，而在於風險能否減緩至不再重要程度。

（一）資料釋出不免須忍受被再識別化之風險，故應建立合理的去識別化標準

由於去識別化、匿名化資料有被再識別化之風險，法律社群乃有「資料效用與隱私保護不易相容，而能兼顧二者之選項有限」印象[351]，因此，去識別化、匿名化成效飽受質疑；對於其未來，如前揭，學界與實務界皆存在激烈辯論，不同見解也將導出分歧規範取向。例如，（1）Paul Ohm 的關鍵主張[352]，認為匿名化已失效而不應將其視為隱私保護的唯一手段，且科技進步漸易由準識別符號（quasi-identifiers）再識別化，個資範圍膨脹而與非個資界線日趨模糊，乃有倡議[353]應揚棄隱私規範對於「識別化」與「去識別化」之區分。（2）Jane (Yakowitz) Bambauer 則持相反觀點[354]，匿名化固有侷限，Paul Ohm 顯然高估危害的風險性，資料流通的益處仍高於再識別化風險；乃另有建議，去識別化、匿名化所提供的保護雖非完美無缺且非足以防止再

---

351 Andrew Chin & Anne Klinefelter, *Differential Privacy As A Response to the Reidentification Threat: The Facebook Advertiser Case Study*, 90 N.C. L. REV. 1417, 1417 (2012).

352 Ohm, *supra* note 120, at 1742-43; Brasher, *supra* note 10, at 231.

353 Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and A New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1876 (2011); Paul M. Schwartz & Daniel J. Solove, *Reconciling Personal Information in the United States and European Union*, 102 CAL. L. REV. 877, 904-16 (2014).

354 Yakowitz, *supra* note 7, at 4.

識別化，但仍可防止缺乏動機、資源與技術的人進行再識別化，故去識別化概念不應輕易揚棄[355]。（3）有別於當前欲追求完美匿名化而以產出結果為基礎（output-based）取向論成敗，Ira Rubinstein 與 Woodrow Hartzog 則提出以程序為基礎（process-based）取向的折衷主張[356]，認為應著重於去識別化、匿名化之過程，而非追求難以企及的安全保證狀態；因此，應著重於風險評估與實施去識別化、匿名化相關技術、安全維護措施而管控資料流通[357]。對此，本文認為資料效用與隱私保護之間，Ira Rubinstein 與 Woodrow Hartzog 之折衷主張應較能兼顧二者需求而務實可行。

1. 資料釋出應承受風險而著重於風險的控管，並非追求完美的去識別化

資安領域早已認知「沒有完美資訊安全」，應跳脫追求完美的迷思，而採行更務實策略。為尋找正確可行的因應之道，不應讓完美成為良善的敵人（the perfect cannot be the enemy of the good）。由前揭再識別化案例可知，追求完美的去識別化（匿名化）也是一種迷思；Ira S. Rubinstein 與 Woodrow Hartzog[358]乃主張，去識別化、匿名化應界定為最小化風險之一種過程，而非難以企及的安全保證之一種狀態。否則，倘要求完全零風險，而不惜一切避開不好的風險，將帶來另一種風險，即無法及時承擔那些可能連帶帶來益處之好的風險[359]。因此，對於資料釋出應採行「承受、忍受風險」（tolerant of risk[360]）取向，將有助於克服關於能否完全（美）、徹底匿名化之爭論。Jane

---

355 Salvatore J. Russo, *Is De-Identification of Personal Health Information in the Age of Artificial Intelligence A Reality or A Noble Myth?*, 22 J. HEALTH CARE COMPLIANCE 55, 58 (2020).

356 Rubinstein & Hartzog, *supra* note 69, at 706, 737.

357 Brasher, *supra* note 10, at 231-33.

358 Rubinstein & Hartzog, *supra* note 69, at 736-37; Hartzog & Rubinstein, *supra* note 118, at 22-24.

359 Michele Wucker（著），許恬寧（譯）（2022），《找出生活中的灰犀牛：認識你的風險指紋，化危機為轉機》，頁60，天下文化。

360 RAPHAËL GELLERT, THE RISK-BASED APPROACH TO DATA PROTECTION 241-242 (2020); Yakowitz, *supra* note 7, at 39-40.

Yakowitz、Felix T. Wu 及 Paul Ohm 等學者亦均已認知，不得不承受風險[361]。質言之，去識別化（匿名化）的論辯應著重於風險的控管，而非完美的追求。

當前欲運用快速發展 AI 與大數據而獲益，如仍要擁有絕對的隱私保護，無疑將困難重重。因此，不宜將絕對的保證視為必然，而應努力建立可合理保護個資的標準[362]。Ira S. Rubinstein 與 Woodrow Hartzog 乃指出，資料釋出之法律與政策，不應侷限於被指控為是否失去效用的去識別化（匿名化）之爭辯，反而應著重於「極小化風險之過程」。上開轉變將可重塑政策論辯之重心，由「完美的去識別化（匿名化）」遠離，而務實導向「風險管理」之過程。再者，因應之道尤應具有「風險忍受、寬容性」，亦即資料釋出政策應著重於「程序」（process）而非「結果」（output），因此，政策目標應用以提升「倘被再識別化之風險、代價」至可承受的程度，而毋庸確保須為「完美的去識別化（匿名化）」。申言之，較為永續可行的政策取向應著重於資料保護所必要之前提要件與程序，故應確保能遵循適當合理程序（包括結合去識別化技術、法律與行政措施）[363]而最小化風險。

2. 倘要求再識別化風險為零，將更難獲取研究資料而不利公益

如前揭，去識別化（匿名化）資料仍有再識別化之風險，當風險愈高，愈會侵害隱私[364]。因此，針對其風險差異，如下列圖表一所示，在去識別化標準的光譜中，主管機關得進行去識別化政策選擇。其中，橫軸乃去識別化標準，100%即指最嚴格標準；縱軸則係再識別化風險；中間直線，則顯示，在某一去識別化標準下可能發生之再識別化風險。當去識別化標準等於「零」時（即屬完全可識別化資料），則再識別化風險為 100%。其實，當再識別化或識別化能力為 100%時，縱採取其他行政安全維護措施以保護個資，仍不存在去識別化。另因行政安全維護措施（如限制員工近用資料）未在技術

---

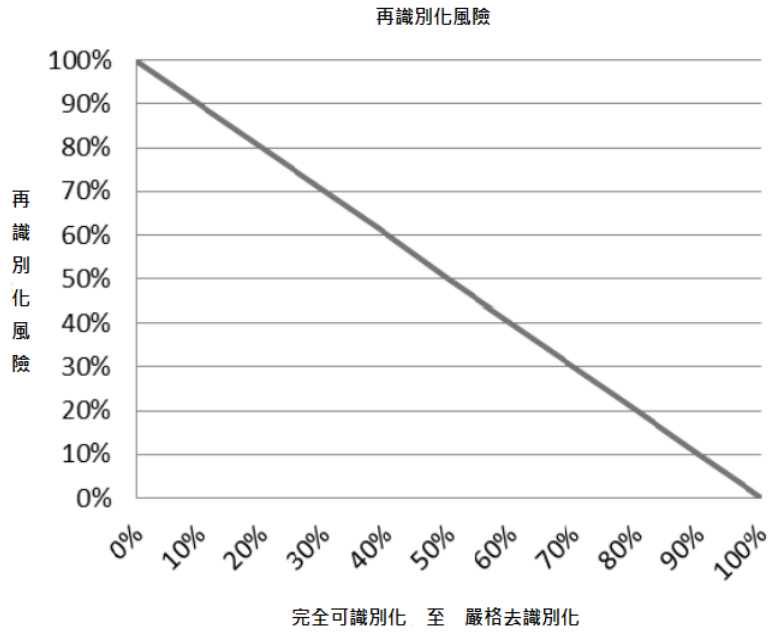
361 Ohm, *supra* note 120, at 1717-18; Wu, *supra* note 280, at 1152; Yakowitz, *supra* note 7, at 2-3; Rubinstein & Hartzog, *supra* note 69, at 736.

362 Russo, *supra* note 355, at 58.

363 Rubinstein & Hartzog, *supra* note 69, at 729-31, 737; Brasher, *supra* note 10, at 233.

364 EL EMAM, *supra* note 41, at 153.

上變更資料本身而僅保護免於被濫用，故這些措施未足以被視為「去識別化」[365]。



【圖一】再識別化風險[366]

※ 資料來源：Yianni Lagos, Taking the Personal Out of Data: Making Sense of De-Identification, 48 Ind. L. Rev. 187, 202 (2014).

去識別化標準寬嚴與再識別化風險高低，呈現反比關係[367]；隨著去識別化標準提高，再識別化風險亦隨著線性下降。惟光譜另一端，完全去識別化是不可行的。倘資料在技術上連結至某個人的機率為零，則將不存在隱私風險；但一個有用處的資料集重新連結至當事人之機率絕非是零。因此，法律不應採行完全不可連結（perfect unlinkability）的不可行標準[368]。去識別化致

365 Yianni Lagos, *Taking the Personal Out of Data: Making Sense of De-Identification*, 48 IND. L. REV. 187, 191, 202 (2014).

366 *Id*. at 202.

367 EL EMAM & ARBUCKLE, *supra* note 25, at 9.

368 Lagos, *supra* note 365, at 190-91, 202.

力於維護資料機密性，同時亦導致資訊流失（information loss）[369]；當個人特徵是資料分析要素時，更減損、限縮資料實用性。

不可能揭露資料同時，又能擔保再識別化風險為零。倘日常生活要求零風險，無異休想踏出家門一步[370]。倘要符合再識別化風險為零之要求，唯一方法即完全不要揭露任何資料；倘揭露任何資料，再識別化機率不會是零[371]。因此，可能被再識別化之風險，應視為整體隱私、個資保護不可避免之一環[372]；否則，倘要求再識別化風險為零，將更難獲取可供研究資料[373]而不利於公益。可合理期待者，乃經去識別化而能利用資料，又能讓再識別化風險降至非常微小[374]。

3. 建立合理的去識別化標準，以在隱私保護與資料效用之間取得平衡

完美的資料釋出政策，應能在多重價值（如法律與科技、資料管制與利用）之間審慎尋求均衡[375]，以在保有資料效用時仍可兼顧隱私保護。資料的可用性與去識別化（匿名化）的程度之間，通常存在一種抵換關係[376]。資料去識別化後，隱私侵害程度通常可大幅降低[377]；惟為保護隱私而去除過多資料將降低資料的科學與研究價值，因此，在保護隱私與保留相關資料以確保資料用途之間，必須有細緻的抵換安排[378]。而為兼顧資料效用，資料轉換（transformation）的程度必須能降低可識別性至適當的可容忍程度，用以合

369 EL EMAM & ARBUCKLE, *supra* note 25, at 35-37.
370 *Id*. at 22.
371 EL EMAM, *supra* note 41, at 135.
372 Lagos, *supra* note 365, at 202.
373 Waldo, Lin, & Millett, *supra* note 50, at 221.
374 EL EMAM & ARBUCKLE, *supra* note 25, at 22-23.
375 Rubinstein & Hartzog, *supra* note 69, at 756.
376 Groos & Veen, *supra* note 101, at 504.
377 Lagos, *supra* note 365, at 199.
378 Katherine Drabiak, *Caveat Emptor: How the Intersection of Big Data and Consumer Genomics Exponentially Increases Informational Privacy Risks*, 27 HEALTH MATRIX 143, 167 (2017).

理確保「資料不可識別的」，但再識別化機率絕非為零，因此，乃衍生可接受風險的門檻、檻值（門檻值）（threshold）（即可接受風險的臨界值）為何之問題[379]。如前揭，倘要揭露資料，要求零風險乃不切實際的；反而應衡量為了追求公益或某些目標而帶給個人的風險[380]是否必要及成比例的[381]，因而設定一個可接受的再識別化機率，而如下述相關立法例乃要求去識別化資料須使個人不可合理地被識別出來[382]。因此，倘實際的再識別化機率在可接受的數值之下，則可進行資料集的揭露、釋出；而風險程度被認為太高時，則拒絕資料的釋出。惟對於任何去識別化方案之難題、挑戰，即可接受再識別化機率風險檻值之設定；其實，並非再識別化機率必須達到零，而是非常微小或可接受的低（acceptably low）[383]。

進而言之，選擇何種風險檻值，應視隱私侵害可能性而定，即應考量資料敏感性與資料原始蒐集時所存在的同意機制。倘隱私侵害可能性被視為較高時，應選擇較低（即較嚴格）風險檻值；而倘隱私侵害可能性被視為較低時，應選擇較高（即較不嚴格）風險檻值[384]。換言之，所應採行去識別化標準之寬嚴，仍應以風險為基礎取向進行評估，按資料釋出之異動風險

379 ARBUCKLE & EL EMAM, *supra* note 5, at 58.

380 類似於前揭之再識別化風險的情境脈絡評估因素，另有主張，可接受的風險應考量因素，如識別技術、資料揭露對於資料主體隱私侵害程度(須視資料敏感性、不當揭露的可能侵害、與資料主體同意的性質與範圍)、潛在攻擊者的動機與能力。Rubinstein & Hartzog, *supra* note 69, at 760. 另NIST指出，有七個項目變數得用以評估風險：資料數量、資料敏感性、資料接收者類型、資料利用、資料處理技術、資料近用控制、同意與消費者期待。NIST, GUIDE FOR CONDUCTING RISK ASSESSMENTS (Sept. 2012), https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf (last visited Oct. 30, 2022); NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS, *supra* note 37, at 10.

381 VICTORIA STATE GOVERNMENT, *supra* note 251, at 18.

382 Kevin C. Gilligan, *Protecting Consumers and Regulating Data: The Need for Comprehensive Federal Oversight of the Direct-to-Consumer Genetic Testing Industry*, 14 DREXEL L. REV. 207, 224-25 (2022).

383 EL EMAM, *supra* note 41, at 135-136 ; Borgesius et al., *supra* note 22, at 2123.

384 ARBUCKLE & EL EMAM, *supra* note 5, at 58.

（transaction risk）高低而定。同一資料集應按異動風險之差異，進行不同程度之去識別化。異動風險較高時（如公開釋出資料、接收者不特定、直接可識別資料、敏感性資料），應採行較嚴格、甚至最高或接近完美的去識別化標準；異動風險較低時（如資料近用受到限制、接收者特定、間接可識別資料、非敏感性資料），則採行較不嚴格的去識別化標準，即足以保護資料。其中，有認為，相較於釋出予學術研究人員，釋出予產業業者的異動風險較高，因可能較難以監督其將如何處理資料[385]。可知，風險檻值的選擇乃根據資料釋出的情境脈絡而定。例如，最為開放的情境脈絡，乃由政府或統計單位所公開釋出的資料。當情境脈絡越開放，資料細微程度應越低；倘公開釋出而對於資料如何利用未加以限制或制衡，則風險最大，故應選擇最嚴格檻值而要符合最高程度的去識別化（匿名化）[386]。乃有主張，前揭 HIPAA 安全港標準，倘資料係供可公開取得者，將最具實用性；但對於研究而言，則可能過於嚴苛[387]。

因此，曾有供研究目的而揭露民眾資料（如癌症登記資料[388]）之情形，被認為 5%與 20%的再識別化比例、機率，係可接受的風險門檻、檻值；因為當資料接收者是較受信賴（如內部或附屬的研究人員）時，則可接受較高的檻值（如 20%）；而當接收者是與資料控管者現在不存在持續性關係的外部研究人員時，則需較低的檻值（如 5%）。按此準則，資料控管者得據以決定資料集概括化（generalization）的程度；以及倘為長期性資料集，得據以決定在風險升高到不能接受前，要揭露多少年度資料[389]。而針對健康資料

---

385 EL EMAM, *supra* note 41, at 4-5, 7; EL EMAM & ARBUCKLE, *supra* note 25, at 38-39, 118-119; Borgesius et al., *supra* note 22, at 2123.

386 EUROPEAN MEDICINES AGENCY, *supra* note 156, at 49; Groos & Veen, *supra* note 101, at 505.

387 El Emam et al., *supra* note 145, at 7.

388 癌症登記資料包含大量可識別的健康資料，不慎洩漏的危害風險將很高。惟癌症登記資料亦能提供醫療專業人士高質量的醫療資料而造福公眾健康。其正是兼具健康隱私益處與代價之例證。Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125, 1153 (2015).

389 El Emam et al., *supra* note 145, at 11; 臺灣類似之風險評估及重新識別機率驗證結

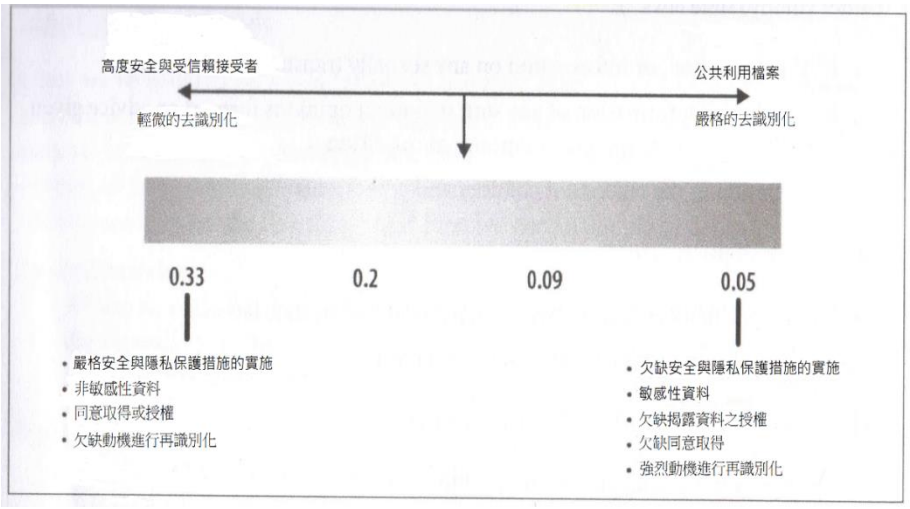
的公開釋出，歐盟藥品管理局（下稱 EMA）則建議應將再識別化機率檻值設定在 0.09（9%）的保守程度。EMA 進一步指出，進行匿名化時，應設定可接受的再識別化風險檻值，包括評估現有的風險減緩措施、某一特定揭露對於試驗參與者隱私侵害的程度、攻擊者再識別化資料的動機與能力。一旦檻值確定，即可衡量再識別化的確實機率；倘再識別化機率高於檻值，則必須進行資料的匿名化。否則，資料的再識別化風險將被視為非常微小（very small）且已達完全匿名化[390]。

另外，有根據過去數十年的先例，而確立資料釋出所可接受之再識別化風險機率；如下列圖表二所示這些先例的數值近年未改變而仍被廣泛使用。其實，這些數值乃受美國、加拿大主管機關與法院所建議。如先例所示，當資料公開釋出，再識別化機率之最大風險檻值將界於 0.09 至 0.05（9%-5%）之區間；當資料釋出予高度安全與受信賴接收者，則最大風險檻值將界於 0.33 至 0.2（33%-20%）之區間。當設定最大風險檻值界於 0.09 至 0.05（9%-5%）之區間時，必須考量到資料的敏感性與資料原始蒐集時已有的同意機制。例如，倘資料具高度敏感性時，將選擇區間較低的檻值；另一方面，倘當事人（病人）明瞭風險而仍明示同意資料公開釋出時，則可在區間擇定較高的檻值[391]。

果，參王興娟（2018），〈行政院主計總處個資去識別化作業辦理情形與成效〉，《主計月刊》，753 期，頁 75-76，http://www.bas-association.org.tw/catalog/arts/010709072.pdf （最後瀏覽日：04/30/2022）。

390 EUROPEAN MEDICINES AGENCY, *supra* note 156, at 49; 憲法法庭111年憲判字第13號判決之吳全峰副研究員意見書，頁4-5。

391 EL EMAM & ARBUCKLE, *supra* note 25, at 38-40.



【圖二】不同的最大風險檻值[392]

※ 資料來源：Khaled El Emam & Luk Arbuckle, Anonymizing Health Data 39 (December 2013).

其實，任何去識別化之標準，應在「降低再識別化風險所帶來的額外隱私保護」與「業者（資料控管者）拒絕去除（scrub）資料所帶來的隱私減損」之間，進行平衡。業者去除資料之比例、程度，乃視效益與成本之運作結果而定。去識別化標準若更加嚴格，須花費更多時間與資源以去除資料。資料效用的減損與執行成本，在嚴格的去識別化標準之下，當成本超過利益時，業者可能放棄任何資料之去除，而繼續使用可識別資料或不再創新；惟資料的分享與利用能創造諸多福祉，應確保其能盡責運作良善。因此，應進行再識別化之風險評估，以建立符合風險比例的去識別化合理、明確[393]標準，讓業者願意採用落實，才能實施久遠，以妥適保護消費者隱私[394]並兼顧資料效用。

---

392 *Id.* at 39.

393 如前揭，就個資、去識別化、匿名化的界定，英國、歐盟、美國法所提出「合理」審查標準，均未要求絕對地「完全排除被再識別化之風險」，亦即排除可能再識別化之絕對主義，而採行較可行的相對主義。

394 Lagos, *supra* note 365, at 200; EL EMAM & ARBUCKLE, *supra* note 25, at 9.

因此，資料釋出之較周全保護政策，應要求業者提供資料的合理保護措施（如實施合理的去識別化技術、最小化資料釋出、維護資安），並針對風險可能程度[395]按產業合理標準而量身打造業者應盡合理責任。據此，除由執法機關（如 FTC）持續監督「業者去識別化之承諾非欺罔」外，如前揭，業者固難以保證「完美的去識別化（匿名化）」，但仍可承諾「已進行資料利用、釋出之風險評估、並按產業標準實施適當安全維護措施」，以利消費者（資料主體）設定正確可行的隱私期待[396]。此外，合理適當的去識別化亦屬一種自我防衛機制，去識別化措施倘符合標準則可做為向執法機關證明「已嚴格善盡對當事人責任」之證據[397]。

（二）立法例亦常採合理識別化、去識別化標準而未要求完全沒有再識別之風險

針對再識別化風險，學者有主張[398]，許多法律其實未要求「個人完全沒有由釋出資料被識別出來之風險」，從過去到現在均呈現出一種風險效用分析（risk-utility analysis）之保守立場[399]。例如，英國資料保護法（Data Protection Act 1998，下稱 DPA）未要求「去識別化、匿名化應完全沒有風險」，而是須減緩再識別化風險至極低之程度；能 100%匿名化最為理想，有時也有可能，但那不是 DPA 所要求的判斷標準。然而，倘再識別化風險具「合理可

395 憲法法庭111年憲判字第13號判決黃昭元大法官部分不同意見書第16段指出，「個資法規定並未進一步針對不同風險程度之目的外利用情形，要求或區別不同的去識別化方法，致健保資料之主管機關享有相當大的選擇空間，而有侵害人民資訊隱私權之風險。」

396 Rubinstein & Hartzog, *supra* note 69, at 707,731-32, 753, 736-37; Hartzog & Rubinstein, *supra* note 118, at 24.

397 EL EMAM & ARBUCKLE, *supra* note 25, at 1-2.

398 Douglas J. Sylvester & Sharon Lohr, *Counting on Confidentiality: Legal and Statistical Approaches to Federal Privacy Law after the USA Patriot Act*, 2005 WIS. L. REV. 1033, 1113 (2005).

399 Yakowitz, *supra* note 7, at 65.

能」時，該資料應被視為個資[400]。*Department of Health* 案[401]高等法院乃指出，倘識別風險高於「不太可能（remote）與合理可能」，則應定性為個資。

按此「風險忍受」概念，去識別化、匿名化的關鍵，應非在於完全排除再識別化風險，而在於能否減緩至不再重要程度。換言之，應著重於減緩風險至再識別化可能性極低之程度[402]，故去識別化、匿名化的方法、方式或標準不在於應確保不可還原的或絕對不可能被再識別化，而在於其風險效用與合理性之評估。

1. 英國、歐盟、美國立法例

類似於上開「風險忍受」概念，國際立法例（如英國、歐盟、美國）普遍採用「合理」識別化[403]、去識別化標準，均未要求絕對地「完全排除被再識別化之風險」，而仍有承擔被再識別化之可能，惟其風險屬於相對較低程度而已，因此，難以確保去識別化、匿名化乃絕對不可還原的。

如前揭，按英國 ICO 所提出，匿名化，乃指在考量資料控管者或其他人所得用以直接或間接識別特定當事人之所有「合理可能」方法、手段後，而仍未能從資料本身或與其他資料相結合而識別特定當事人。而按歐盟 GDPR Recital 26，為了確認某些方法、手段得否「合理可能」用以識別特定當事人，必須考量所有客觀情境因素，如識別所需之成本與耗時長短、資料運用與科技發展當時可資利用之技術（如運算能力）與工具。故資料倘有識別的合理風險，應被視為個資。GDPR、ICO 與上開 *Breyer* 案均採風險為基礎的取向，以相對立場而判斷資料是否具個人屬性[404]。WP29 亦指出，經匿

---

400 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 6.

401 The Department of Health v Information Commissioner [2011] EWHC (Admin) 1430.

402 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 16.

403 資料可識別個人的程度，取決於資料接收者身分、對接收者與接觸資料所加諸之契約、安全管控措施（如限制再識別化）等因素。ARBUCKLE & EL EMAM, *supra* note 5, at 7.

404 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 16, 58, 97; Finck & Pallas, *supra* note 94, at 14-15. 其實，關於個資定義，有兩種理論。按絕對主義（又稱客觀說），應考量所有的可能性及機會，只要有人能辨識出資料當事人，即為個資；換言之，只要世界上有任何一人以任何方法可探查當事人身分，即是。

名化的資料仍有潛在識別的可能性，因任何用以使資料匿名化的技術性或組織性措施仍會殘留固有的再識別化風險[405]。ICO 承認，因常未能明確確認何種資料現已或將來可供取用，經由資料連結而再識別化的風險乃不可預測。愛爾蘭資料主管機關亦採類似立場而主張，毋須證明「某一匿名化技術的有效，乃因資料主體不可能被識別出」，反而應證明「按某個案與技術實況，資料主體不可能被識別出，故資料可被視為匿名的」[406]。因此，為了判斷某資料是否達成法律上（legally）匿名化，在某些預期範圍內，評估資料控管者與第三人所有可用以再識別化資料主體之「合理可能」方法、手段即可[407]。

ICO 並提出「有心侵入者測試」（motivated intruder test）標準，即一位侵入者若有心、動機嘗試時，能否成功重新再識別出個資已被匿名化之特定當事人。所謂「有心侵入者」，乃指對於該資料事前一無所知之一個具有理性行為能力（reasonably competent）之人，能夠接觸相關資源（如網路、圖書館、所有公開資料）及運用相關調查技能（如詢問對於特定資料當事人有特別認識之人、或公開徵求擁有相關資訊之人），但其未具有特定專門知識（如電腦駭客技能）、未能接觸專門設備且未能以犯罪手法（如竊盜）而未能取得已被安全保管之資料，仍藉由已被匿名化之個資而意圖重新再識別出特定當事人[408]。類似地，在判斷個案中何種程度的匿名化是必要的，愛爾蘭

按相對主義（又稱主觀說），識別的主體及方法不能毫無限制，主體限於資料管理者或有範圍的主體，方法限於有實際可能的方法，排除僅為假設性的方法。歐盟法院判決、WP29、英國ICO均傾向相對主義。江耀國、黃子宴（2019），〈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》，58期，頁8-14。

405 *Opinion 05/2014 on Anonymisation Techniques* 6-8.

406 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 16, 58, 97; DATA PROTECTION COMMISSION (DPC), GUIDANCE NOTE: GUIDANCE ON ANONYMISATION AND PSEUDONYMISATION 5 (2019), https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf (last visited Apr. 14, 2022); Finck & Pallas, *supra* note 94, at 14-15.

407 Stalla-Bourdillon & Knight, *supra* note 4, at 292.

408 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20, at 22-25. ICO認為，「有心侵入者測試」標準具有實用性，乃因所設定的再識別化風險之折衷標準，高於

亦採某一人（侵入者或內部人）按當前（未來仍應隨時評估）可用技術與資料而得用以識別資料主體的所有「合理可能」方法之標準[409]。

如前揭，按美國 FTC，倘有合理基礎而可相信「某一特定檔案中的殘留資料不能用以識別某一個人」，則應被視為成功的去識別化。按 HIPAA 規定，沒有「合理基礎」而相信「可能用以識別某一個人」，則非屬個人可識別健康資料。按加州、科羅拉多州及維吉尼亞州個資保護法規均指出，去識別化，係不能「合理地」識別、或直接或間接連結至某一個人。

綜上，按上開立法例普遍所採「合理」的識別化、去識別化標準，均未要求絕對地「完全排除被再識別化之風險」；故縱以非「合理可能」方式而有重新再識別化可能而非不可還原的，亦屬於去識別化、匿名化資料供運用而應承擔風險之範圍。據此，有時不得不習慣於一個比預期更少隱私保護的世界。

2. 我國法

對於是否要求「個人完全沒有由釋出資料被識別出來之風險」，我國法未明定，而法院見解分歧。

(1) 個資法未明定

關於去識別化（或匿名化）的定義與方法，我國現行個資法未在任何條文直接使用該等詞彙，僅在某些條文[410]規定「資料經過提供者處理後或經蒐集者依其揭露方式無從識別特定之當事人」而間接闡述之。個資法施行細則第 17 條規定：「……所稱無從識別特定當事人，指個人資料以代碼、匿名、

評估相對非專業（relatively inexpert）大眾，而低於評估具有特殊專業（specialist expertise）、分析能力或事前資訊之人能否達成再識別化的標準。樓一琳、何之行（2017），〈個人資料保護於雲端運算時代之法律爭議初探暨比較法分析：以健保資料為例〉，《臺大法學論叢》，46卷2期，頁405。

409 THE INFORMATION COMMISSIONER'S OFFICE, *supra* note 20; DATA PROTECTION COMMISSION, *supra* note 406, at 8.

410 個資法第6條第1項但書第4款、第16條但書第5款、第19條第1項但書第4款、第20條第1項但書第5款。

隱藏部分資料或其他方式，無從辨識該特定個人者。」對此，常被引用之法務部民國 103 年 11 月 17 日法律字第 10303513040 號函釋：「……個人資料，運用各種技術予以去識別化，而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。」惟對於所謂「無從識別」之定義與方法，是否須確保去識別化（或匿名化）不可還原（逆）或絕對不可能被再識別化？識別是否須以「合理可能」方法為限？則語意不明，有待解釋[411]，可能被質疑不符法律明確性[412]而有違憲之虞；惟憲法法庭 111 年憲判字第 13 號判決則認定「『無從識別特定之當事人』文義，尚非難以理解，且應已足使一般受規範者得預見，並可經由司法審查加以認定及判斷」，「與法律明確性原則、比例原則尚屬無違，不牴觸憲法第 22 條保障人民資訊隱私權之意旨。」

(2) 法院裁判見解分歧

關於「無從識別」的定義與方法，是否須確保去識別化（或匿名化）不可逆、不可還原或絕對不可能被再識別化？識別是否須以「合理可能」方法為限？對此，就全民健康保險研究資料庫運用爭議，我國法院裁判有所探討，行政法院見解分歧，惟憲法法庭已為判決。

---

411 對此，我國經濟部標準檢驗局參考國際標準（2012年版ISO/IEC 29191）所制定 CNS 29191 「資訊技術－安全技術－部分匿名及部分去連結鑑別之要求事項」國家標準，建立個資去識別化之具體要求及控制措施之檢驗標準，雖不具法拘束力，在於提供去個人識別化時得參考之準則及施作流程，並鼓勵民間組織等取得認證，以期能達到開放資料「無從識別特定當事人」的成效。范姜真媺，前揭註134，頁46；財團法人電信技術中心，前揭註127，頁131。

412 例如，依法務部函釋，可推論出：「提供者」可將「可直接識別之個人資料」，提供給「蒐集者」，只要最終符合「蒐集者依其揭露方式無從識別特定之當事人」即可，在此種標準混亂的操作邏輯下，實難想像其侵害人民資訊隱私權程度之嚴重程度會有多高（參憲法法庭111年憲判字第13號判決劉靜怡教授意見書，頁11）。另本文認為，倘個資未經處理（去識別化或匿名化）逕分享予不在原始蒐集目的內之第三人，乃違反個資保護之使用限制（use limitation）原則；揆諸國外立法例，罕見容許此種分享型態。可知，因規定語意不明確，可能出現違反個資保護基本原則之解釋結果。

A. 是否以「不可還原」為要件

例如，臺北高等行政法院 103 年度訴更一字第 120 號判決：「較有利於風險之控管，……，其資料去識別化之程度可相對放寬……可採取『可匿（應為「逆」之誤）之擬匿名化資料』方式進行去識別化……只要原資料保有者並未將對照表或解密方法等連結工具提供給資料使用者，其釋出之資料無法透過該資料與其他公眾可得之資料對照、組合、連結而識別出特定個人時，該釋出之資料即屬無法直接或間接識別之資料而達法律規定去識別化之程度。」可知，該判決認為，擬匿名化資料符合法律規定去識別化之程度。另按該判決所引用法務部報告，所謂「擬匿名化資料」（pseudonymized data）（可分為可逆、不可逆）另有譯為「假名化資料」[413]；而如前揭，國外見解亦有認為「可逆、可還原」的假名化（乃以「分開保管並受制技術安全拘束」的假名代替資料主體身分或識別符號）係某種形式的去識別化。質言之，該判決不以「不可還原」為去識別化之要件。

最高行政法院 106 年度判字第 54 號判決：「……未達成『去識別化』作業應有之實證效用（即『徹底切斷』資料內容與特定主體間之連結），該收受之資料仍具『個人資料』屬性）……」該判決乃採須達「不可還原」的「徹底切斷連結」才符合去識別化的要件。

憲法法庭 111 年憲判字第 13 號判決：系爭規定一（個資法第 6 條第 1 項但書第 4 款）之意旨，「係指……應採取去識別化之措施，使資料不含可直接識別特定當事人之資訊，但其資料仍屬可能間接識別特定當事人之資訊之情形。[414]」、「查個人健保資料包含……高敏感特種個資，具有高度個體差異，於客觀上非無以極端方式還原而間接識別特定當事人之可能性，此為科學上之事實。因此，個人健保資料無論為原始型態或經處理，均必然仍屬『得直接或間接識別該個人』之資料……。[415]」、「或有主張將個人健保資

[413] 林裕嘉（2017），〈公務機關利用去識別化資料之風險評估及法律責任（上）〉，《司法周刊》，1852期，頁2-3。

[414] 參本號判決理由第40段。

[415] 參本號判決理由第36段。

料處理成為完全不具還原識別可能性之匿名資料再予利用，同樣能達成系爭規定一之法定目的等語。……匿名資料固非全然不具學術研究價值，但已喪失病歷、醫療、基因及健康檢查資料作為學術研究樣本時可擇定變因交互比對、建立相關性之特性者，將無從達成系爭規定一所欲追求之特別重要公益目的。是系爭規定一以去識別化……為合法蒐用要件，屬最小侵害手段。[416]」可知，該規定所要求「去識別化之敏感個資，係指非可直接識別，但經由極端方式仍有還原可能之個資」[417]。對此，本判決因僅要求資料經去識別化處理即符合要求，而不要求須為完全不具還原識別可能性之匿名資料[418]，可謂乃採取較有利於資料多元分享運用之解釋取向；另一方面，如本判決及本文前揭，客觀上仍非無以極端方式還原而間接識別當事人之可能，且匿名資料所導致的資訊流失若過度而喪失可交互比對、建立相關性之特性者，將無助於公益之達成。因此，如本文前揭，資料釋出不免須忍受被再識別化之風險，故對於所應採取去識別化措施、標準之要求，亦應考量當時合理技術與成本，而排除耗費不成比例成本與時間才能達成還原之情形（如下述 B 單元）。

綜上，對於去識別化（匿名化）後資料是否須達「不可還原」才符合去識別化（或匿名化）定義的要件，行政法院見解不一；憲法法庭則認為去識別化後資料仍有還原可能性，而似認為「匿名資料為完全不具還原識別可能性」（但仍有持部分不同意見者[419]）。惟如本文前揭，當前常有眾多資料可

---

416 參本號判決理由第56段。

417 參本號判決楊惠欽大法官部分不同意見書，頁9。

418 然而，針對本號判決多數意見，黃昭元大法官部分不同意見書第20段指出：「系爭個資法規定之去識別化措施仍非最小侵害手段」、「……，系爭個資法規定就去識別化程度之要求，顯然過於寬鬆。為保障人民之資訊隱私權，最完整的方式本來應該是匿名化，斷開資料與主體間的連結，使之無法或極難以再還原連結，如此個資也就不再是個資法所保護的資料。如果無法匿名化處理，至少也應明定更多的刪除項目，以增加還原連結的難度，並以此為對外提供利用的基本資料組資訊，而不是如目前之僅刪除姓名及住址二個變項資料。……」可知，此處「匿名化，斷開資料與主體間的連結，使之無法或『極難以』再還原連結」、「以增加還原連結的難度」之意見，似未持「匿名資料為完全不具還原識別可能性」之見解。

419 同前註。

供交叉比對，不論去識別化或匿名化，不可避免均有被再識別化風險，已難以維持不可還原而確保絕對不可能被再識別化之狀態，故本文認為，除非亦採行類似於本文前揭「個資的匿名化，應以考量合理可能使用之所有手段為限（即排除須投入不成比例之不合理手段），而使其不可還原地不可能識別資料主體」（詳見本文「貳之三之（二）之 2 之（1）之 B」單元及下揭單元 B）之解釋、界定方式外，否則，不宜以「不可還原」為去識別化、匿名化或「無從識別」定義要件。

B. 是否以「合理可能」方法為限

至於「去識別化」的方法、技術，是否如上開國際立法例普遍採用以「合理可能」方法、手段等為限，國內法院裁判見解不一。

例如，臺北高等行政法院 103 年度訴更一字第 120 號判決：「控管去識別化程度，應進行整體風險影響評估，綜合考量個人資料類型、敏感性程度、對外提供資料之方式、引發他人重新識別之意圖等因素，判斷去識別化之技術類型或程度……即資料經過編碼方式加密處理後，處理後之編碼資料已無從直接或間接識別特定之個人……」可知，該判決對於去識別化方法、技術未明示應以「合理可能」方法、手段等為限。

最高行政法院 106 年度判字第 54 號判決：「……刻意鎖定特定主體，主動搜尋與該主體身分有關之各式資訊，再與『個人資料』作連結之方式，此等該特定主體即使有隱私權受到侵犯，……是一開始連續不當之私人資訊探究行為。因此此等結果之防止，已不在『去識別化效用強度』之法定標準範圍內。」從該判決可知，對於「無從識別特定之當事人」之判斷，進行識別的主體及方法並非毫無限制，而非完全不考慮其「合理可能」性。

憲法法庭 111 年憲判字第 13 號判決：「其提供者至遲於揭露時須為已經採取去識別化措施處理，……。[420]」、「系爭規定……課予採取去識別化措施之義務，使一般人採取當時存在技術與合理成本，在不使用額外資訊時，不能識別特定當事人。雖個人健保資料於客觀上非無以極端方式還原而間接

[420] 參本號判決理由第53段。

識別特定當事人之可能性，惟……所採之去識別化手段已足大幅降低蒐用個人健保資料所生之個人資訊隱私權所生之侵害。[421]」可知，對於「無從識別特定之當事人」之判斷，並不要求確保去識別化不可還原或絕對不可能被再識別化，進行識別的方法倘採取「當時存在技術與合理成本」之「合理可能」措施，即符合法律課予所應善盡去識別化之義務。類似地，謝銘洋大法官部分不同意見書[422]亦指出，「『匿名』……其雖未必達到絕對無法還原的程度，但至少要達到依資料處理時的科技發展程度無法還原，或縱使可以還原但所耗費的成本與時間不成比例的程度（GDPR Recitals 26 參照）。」可知，其亦以採用當時科技所耗費的成本與時間「合理可能」之方法、手段，為判斷是否已達去識別化、匿名化程度之標準。

上開最高行政法院 106 年度判字第 54 號判決及憲法法庭 111 年憲判字第 13 號判決之判決用語或與前揭國際立法例普遍採用以「合理可能」方法、手段等為限之合理標準、用語不盡然相同，但相同者似皆在排除以不擇手段（即不考慮識別所需投入成本、時間、技術之合理性）進行再識別化而質疑「去識別化」（或「匿名化」）成效不彰之情形，以在隱私保護與資料效用衝突間謀求平衡並兼顧二者運作之實際需求。

（三）小結

大數據時代，已被去識別化、匿名化資料仍可能與其他資料相結合，而有再識別化之可能。對此，關鍵之處，應非在於完全排除被再識別化之風險，而是在於被再識別化風險能否被減緩至不再重要之程度。因此，資料釋出不免須忍受被再識別化風險，否則，倘要求再識別化風險為零，將更難獲取可供研究資料而不利於公益。

類似於上開「風險忍受」概念，國際立法例普遍採用以「合理可能」方法、手段等為限之「合理」識別化、去識別化標準，均未要求絕對地「完全排除被再識別化之風險」，而仍有承擔被再識別化之可能，惟其風險屬於相

421 參本號判決理由第54段。

422 參本號判決謝銘洋大法官部分不同意見書，頁6。

對較低程度而已，難以確保去識別化、匿名化乃絕對不可還原的。對此，我國法雖未明定，行政法院曾見解分歧，但憲法法庭 111 年憲判字第 13 號判決亦已趨向採類似的合理標準而不要求確保絕對不可逆、不可還原。

二、去識別化應兼採技術、行政與法律措施而降低再識別化風險

（一）資料越具識別性，安全措施應更嚴謹

去識別化架構之成敗，最終將視用以平衡「資料預定目的」與「減緩再識別化風險的安全維護與控管措施」之成效而定[423]。某些情形，毋庸進一步限制，即可分享去識別化資料；有些情形，則需額外安全維護措施才可釋出[424]。由於去識別化未能使得被再識別化風險下降至零，常須納入額外安全維護措施，以因應殘餘風險。例如，有主張，由於要達成適當、有效的匿名化是不容易的，故有時，匿名化資料集的釋出必須經授權方式為之，不得完全以資訊公開方式為之；去匿名化而被再識別化的風險越高，更有理由要求應以授權方式為之。而授權機制應要求，當被授權人察覺個人可能或已經被再識別化時，應通知授權人[425]。

而在判斷應採取何種管制措施以確保不同程度可識別性資料集之可管控隱私風險，應視具體個案情況而定[426]，並無千篇一律標準。進一步言，資料安全與相關政策乃視情境脈絡的敏感性而定；具體上，所應採取安全維護措施之寬嚴程度，常須視資料的敏感性、業務營運規模與性質及所面臨風險類型而定[427]。資料越具識別性或敏感性，隱私安全管制措施應更嚴謹；反之亦然。具體上，所應採行去識別化程度應與再識別化風險程度成比例關係；釋出資料之再識別化風險越高，去識別化程度應越高。

---

423 Polonetsky et al., *supra* note 8, at 620-21.

424 GARFINKEL, *supra* note 5, at 38.

425 McGraw & Leiter, *supra* note 179, at 443; Borgesius et al., *supra* note 22, at 2124.

426 Polonetsky et al., *supra* note 8, at 620-21.

427 Rubinstein & Hartzog, *supra* note 69, at 734.

資料釋出模式亦將左右去識別化程度之要求；相較於公開釋出資料而要求較高程度之去識別化，非公開資料較不供取用而有較高安全保護，則隱私風險較低而去識別化程度要求亦較低[428]。此外，針對所欲採行去識別化之技術、安全與控管措施，應進行風險評估並記錄決策考量理由[429]，以善盡個資保護之責。

（二）去識別化應考量技術、行政及法律措施

倘欲分享、利用去識別化資料，應結合數種安全維護措施、方法，以降低、最小化再識別化風險[430]。過去的因應之道幾乎僅關注技術性課題（以數學與統計學為主），再識別化案例固未必具代表性，但其容易化[431]乃不再迴避過去認為難以加以量化的麻煩人性考量，而亦從新的社會學、心理學與組織性角度著手因應。技術進步固有助於解決問題，仍未能完全取代政策的功能[432]。惟單靠政策也不足以充分保護隱私；技術的控制措施雖非隱私規範的替代品，若妥善運用，仍有助於貫徹政策目標[433]。乃有主張，去識別化應理解為將技術及行政（含法律）措施均納入考量的一種過程[434]，以確保隱私。其實，比起單靠較嚴格技術措施本身，合理良善的技術與行政措施之結合，可帶來更低的再識別化風險[435]。簡言之，去識別化的進行應先按再識別化風險評估，而兼採符合比例之合理技術、行政與法律措施。GDPR 第 24（1）、

---

428 INFORMATION AND PRIVACY COMMISSIONER OF ONTARIO, DE-IDENTIFICATION GUIDELINES FOR STRUCTURED DATA 7-9 (2016), https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf (last visited Apr. 14, 2022); EL EMAM & ARBUCKLE, *supra* note 25, at 38-39.

429 Polonetsky et al., *supra* note 8, at 621.

430 GARFINKEL, *supra* note 5, at 38; Rubinstein & Hartzog, *supra* note 69, at 729.

431 CENTER FOR DEMOCRACY & TECHNOLOGY, *supra* note 175, at 8.

432 Ohm, *supra* note 120, at 1751, 1761.

433 CENTER FOR DEMOCRACY & TECHNOLOGY, *supra* note 175, at 13.

434 參GDPR第4(5)條; FEDERAL TRADE COMMISSION, *supra* note 75, at 22; Achatz & Hubbard, *supra* note 70, at 1-9.

435 Yianni Lagos & Jules Polonetsky, *Public vs. Nonpublic Data: The Benefits of Administrative Controls*, 66 STAN. L. REV. ONLINE 103, 108 (2013).

32（1）條即規定，資料控管者必須根據風險而採行適當的技術與組織性措施[436]。

資料接收者再識別化的能力與動機（如能否因此獲利）越大，則再識別化風險將越大[437]。行政、法律措施雖未在技術上改變資料，其本身也不能算是去識別化，卻有助於控管不可預見風險（unforeseen risk）而保護資料免遭濫用；其措施包括政策、契約、訓練等[438]，如以資料利用契約限制資料接收者行為而禁止去識別化資料連結至其他資料或與他人分享而進行再識別化[439]、員工近用資料的限制[440]、隱私衝擊評估、定期查核、最敏感資料禁止釋出、資料外洩通知、損害賠償責任、安全維護等措施，均有益整體風險控管[441]。另對於會接觸去識別化資料的人，應教育訓練以組織（行政）與技術措施強化隱私安全保護[442]。我國法院及實務界[443]均亦肯認，行政（法律）管制措施（如契約）對於禁止再識別化之重要性。

---

436 風險（risk）在GDPR規定中，重複了75次；可知，GDPR規範架構乃建立在以風險為基礎取向（risk-based approach）之上。Kloc et al., *supra* note 83, at 7.

437 EL EMAM, *supra* note 41, at 153.

438 Polonetsky et al., *supra* note 8, at 620-21.

439 GARFINKEL, *supra* note 5, at 38;如前揭，加州隱私權法、科羅拉多州隱私法及維吉尼亞州消費者資料保護法關於去識別化定義之相關規定均要求，持有去識別化資料之業者應公開承諾不會嘗試再識別化該資料，且以契約課予資料接收者應遵守相關要求。

440 Lagos, *supra* note 365, at 191; Polonetsky et al., *supra* note 8, at 620-21.

441 EL EMAM, *supra* note 41, at 7, 153; BROWN ET AL., *supra note* 248, at 31-32.

442 CENTER FOR DEMOCRACY & TECHNOLOGY, *supra* note 175, at 15.

443 例如，就全民健康保險研究資料庫運用爭議，臺北高等行政法院103年度訴更一字第120號判決肯認：「得與資料使用者約定禁止重新識別資料之義務及其他資料利用之限制等，較有利於風險之控管，……，其資料去識別化之程度可相對放寬，可提供含有個體性、敏感性之擬匿名化資料」。又如，交通部所訂定「交通數據流通管理作業要點」即禁止重新識別含去識別化個資屬性資料。

（三）同時採用技術及行政（含法律）措施，可建構兩層獨立的強化保護機制

同時採用技術及行政（含法律）管控措施之優勢，在於得以建構兩層獨立的強化保護機制。所謂「獨立」，係指「違反行政安全維護措施的機率」（行政性風險） 與「不當再識別化資料的機率」（技術性風險）彼此不具連動性。倘行政性風險乃獨立於技術性風險之外，則兼採技術與行政管控措施將可劇烈降低整體隱私風險。假設行政性風險為 1%而技術性風險亦為 1%，則隱私侵害機率下降至 0.01%。惟資料分享予其他業者時，此雙重保護優勢將下降，因每一業者內部均有一獨立的資料侵害行政性風險，尤其，分享業者眾多時，優勢甚至可能完全消失[444]。

縱同時有技術及行政（法律）管控措施，但實際上業者仍常可按自身動機而決定要保護或濫用資料，因而對於再識別化行為可能難以偵測；乃有認為，禁止再識別化注定要失靈。因此，「信任」在去識別化機制、管控措施上仍扮演關鍵角色，沒有隱私保護不會涉及對於資料蒐集者的某程度信任、信賴關係；社會有時仍不得不信任業者會以適當方式而利用資料，以利創新[445]。

然而，資料蒐集者仍應持續藉由技術、行政與法律措施，降低資料之個人可識別性，以在促進重要利用與有價值的研究時仍可保護隱私[446]。

（四）小結

去識別化程度應與再識別化風險程度成正比，再識別化風險越高，去識別化程度應越高。倘欲分享利用去識別化資料，則應結合數種措施、方法，以降低再識別化風險。乃有主張，去識別化應理解為將技術及行政（法律）措施均納入考量之過程；合理良善技術與行政措施的結合，可帶來更低的再識別化風險。同時兼採技術及行政（法律）措施之優勢，在於得以建構兩層

444 Lagos, *supra* note 365, at 201.

445 *Id*. at 198.

446 Polonetsky et al., *supra* note 8, at 594.

獨立的強化保護機制。惟去識別化的進行應先按再識別化風險評估，而兼採符合比例之合理技術、行政與法律措施。

三、課予民刑事責任而禁止不當再識別化

某些立法（如 HIPPA）假定「去識別化資料能確保完全的匿名化」，故倘資料經去識別化，則不在法律保護範圍，可不受限制地自由散佈。去識別化（匿名化）程序縱符合法律要求，惟資料無論如何（如以未能合理預見方式）仍有被再識別化之可能，因此，再識別化行為乃構成對於後來變成個資之不當蒐集，而此種蒐集顯然等同於牴觸本欲進行匿名化者之目的[447]。

由於去識別化仍有可能再識別化，乃有主張應立法禁止不當再識別化，並對於濫用資料者課予民刑事責任，以最小化被再識別化風險[448]。

（一）應立法禁止不當再識別化之主張

某些隱私專家提議，除核准的再識別化研究外，應藉由各種處罰而禁止不當再識別化，尤應立法禁止惡意再識別化，而立法範圍應擴及再識別化風險的衡量、保障參與研究民眾免於隱私風險與提供充分救濟。另當資料已供公開取用或因難以偵測再識別化行為，再識別化禁止令雖可能難以執行，惟較嚴格處罰、提供求償權、明確的近用控制與稽核程序仍有助於克服執行上的困難與偵測。立法禁止不當再識別化，亦有助於資料科學家精確判斷額外的再識別化風險[449]。再者，我國亦有支持法制上外控機制的強化而非只仰賴研究人員的自律，如黃昭元大法官所言[450]：「……。甚至也有人主張：許多

---

447 Mark Phillips, Edward S. Dove & Bartha M. Knoppers, *Criminal Prohibition of Wrongful Re-identification: Legal Solution or Minefield for Big Data?*, 14 BIOETHICAL INQUIRY, 527, 532 (2017), https://link.springer.com/content/pdf/10.1007/s11673-017-9806-9.pdf (last visited Apr. 14, 2022).

448 Ryan Abbott, *Big Data and Pharmacovigilance: Using Health Information Exchanges to Revolutionize Drug Safety*, 99 IOWA L. REV. 225, 255-56 (2013).

449 Ohm, *supra* note 120, at 1758; Ahn, *supra* note 337, at 790-92, 805-06.

450 參憲法法庭111年憲判字第13號判決黃昭元大法官部分不同意見書第25段。

研究者，尤其是從事量化研究者，本於其專業訓練及研究倫理之要求，多半不會特意去還原並再識別所取得個資之主體為誰，……。不過，……，不能只靠主管機關及研究人員的自律，而仍須有適當的外控機制，包括個資主體（潛在被害人）之自我防衛機制。就算實務上至今並未出現廣泛、重大損害的實例，也不等於就當然要棄守法制上的個人最後防衛手段，而只寄望於國家或他人的善意。」

其實，藉由對於資料去識別化，資料控管者表示、傳達「保護資料主體隱私」之意思，資料主體乃信賴該表示而同意提供其資料；進行再識別化之競爭對手乃阻礙該意思並削弱該同意，因此，應立法禁止再識別化行為[451]。加州即少數州之一，制定了比聯邦 HIPAA 門檻更嚴格的健康隱私法規。加州健康資料交換實施原則（Health Information Exchange Practice Principles[452]）規定，除法律明定外，去識別化個人健康資料不可再識別化；去識別化資料被再識別化，即應受法律保護；倘有合理基礎相信「去識別化資料可用以識別某個人」，則不應加以揭露。對此，加州 CCPA 亦明定[453]，被再識別化的資料不再豁免法律拘束，仍應遵守資料隱私及安全相關的聯邦及州法（如 HIPAA、CCPA）。日本亦立法禁止匿名加工資料處理業者將匿名加工資料與其他資料組合、比對，而復原個資[454]。

（二）課予民事責任之主張

學者有建議立法為基礎的契約解決之道，由資料揭露者與接收者締結契約而釐清雙方義務，賦予受侵害資料主體求償權（此種求償權［訴因］乃 HIPAA 所未賦予）；且禁止進行再識別化，否則，將有民刑事責任；以及接

---

451 Ohm, *supra* note 120, at 1758.

452 CAL. CODE REGS. tit. 22, § 126030 (2015).

453 CAL. CIV. CODE § 1798.146.(a)(4)(B); 孫敏超（2021），〈美國加州修正加州消費者隱私法健康資料去識別化相關規定〉，《科技法律透析》，33卷1期，頁14-16。

454 范姜真媺，前揭註134，頁34。

收者應維持技術、行政等維護措施。藉由交互承擔責任，接收者得向揭露者保證「不會因資料移轉而衍生責任」[455]。

由於資料可識別性、可利用性、隱私保護與去識別化程序成本之間，總存在某程度之抵換難題；科技有時固能減緩，但終未能徹底消除此種抵換困境。因此，立法所能做的，乃建立一法律框架，而容許資料揭露者與接收者能自願基於外部可執行條件（如契約責任）而確保隱私[456]。例如，前揭美國 FTC 或科羅拉多州隱私法規定，資料控管者應以契約課予資訊接收者應遵守相關要求（含公開承諾不會嘗試再識別化），以維持去識別化資料的形式。

如前揭，我國執法者及實務界[457]亦有以立法或契約禁止再識別化。

（三）課予刑事責任之主張

1. 支持之主張

某些學者主張，雖再識別化很少發生，但對於再識別化行為入罪化，乃屬適當的因應之道，對外發出「該行為不被容忍」信號。其中，Jane Yakowitz 主張[458]，應擴大責任適用範圍，不必將再識別化之刑責侷限於那些簽約明示承擔責任之人，並應適用於那些從事揭露非公開資料與身分予他人之人。為避免將無辜研究人員也入罪化，應將刑事責任侷限於故意犯。進一步言，對於僅某些特定數量的研究人員可接觸的資料集而言，藉由資料利用契約而承諾不進行再識別化的義務承擔，即足以提供有效的違約監督。惟當資料公開釋出或可自由下載時，則引進刑責有助嚇阻作用[459]；Sejin Ahn 亦指出，因

---

455 Ahn, *supra* note 337, at 790-92; CENTER FOR DEMOCRACY & TECHNOLOGY, *supra* note 175, at 9-10; Robert Gellman, *The Deidentification Dilemma: A Legislative and Contractual Proposal*, 21 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 33, 47-52 (2010).

456 Gellman, *supra* note 455, at 47.

457 例如，以契約禁止全民健康保險研究資料庫資料之重新識別，及以立法禁止交通數據之重新識別。

458 Yakowitz, *supra* note 7, at 48-49.

459 Phillips et al., *supra* note 447, at 533.

刑事責任適用範圍比侷限於資料揭露者與接收者間的契約更為廣泛，毋庸締結契約而大眾任何一人均可能是資料接收者，尤其，公開資料無庸違反安全措施即可取得而可再識別化[460]。Robert Gellman 主張，不管刑事處罰是否適當，課予刑責將有助於防堵惡意競爭對手進行再識別化[461]。Jorge Contreras[462]亦提議，對於基因資訊進行再識別化，課予民刑事責任。

另一項，考量違反個資保護之民事訴訟原告舉證責任的困難，則應該擴大刑事責任的適用範圍[463]。為重拾對於匿名化之信心，英國有主張[464]，對於蓄意與過失再識別化個人之行為，應課予刑責。英國資料保護法（Data Protection Act 2018）第 171（1）條即規定，未經進行個資去識別化之控管者的同意，而明知或重大過失（knowingly or recklessly）再識別化經去識別化個資之資訊，乃構成犯罪。而在臺灣，倘不符個資法規定（如第 15、19 條第 1 項）要件，卻對去識別化資料進行再識別化而蒐集個資，除負民事賠償責任外，亦可能構成刑事責任[465]；而不法的再識別化行為較可能觸犯個資法第 41 條「意圖為自己或第三人不法之利益或損害他人之利益，而違反第六條第一項、第十五條、第十六條、第十九條、第二十條第一項規定，……，足生損害於他人者，處五年以下有期徒刑，得併科新臺幣一百萬元以下罰金。」規定。關於該條所稱「意圖為自己或第三人不法之利益或損害他人之利益」中之「利益」，是否僅限於財產上之利益？實務上有採肯定說而判決無罪或採否定說而判決有罪，致有不同結果，而違公平原則（參照最高法院

460 Ahn, *supra* note 337, at 791.

461 Gellman, *supra* note 455, at 49**;** CENTER FOR DEMOCRACY & TECHNOLOGY, *supra* note 175, at 9-10.

462 Contreras, *supra* note 62, at 46.

463 Phillips et al., *supra* note 447, at 533.

464 NATIONAL DATA GUARDIAN FOR HEALTH AND CARE, REVIEW OF DATA SECURITY, CONSENT AND OPT-OUTS 8 (2016), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF (last visited Apr. 14, 2022).

465 林裕嘉（2017），〈公務機關利用去識別化資料之風險評估及法律責任（下）〉，《司法周刊》，1853期，頁3。

109 年度台上字第 1869 號刑事判決）；對此，按最高法院 109 年度台上大字第 1869 號刑事裁定，個人資料保護法第 41 條所稱「意圖為自己或第三人不法之利益」，應限於財產上之利益；至所稱「損害他人之利益」，則不限於財產上之利益。除此之外，我國刑法亦可仿效上開英國立法例，專門為不法再識別化行為另立一罪，以求罪刑法定之明確性，避免民眾誤蹈法網。

2. 反對之主張

對於再識別化行為入罪化，卻面對難以執行之挑戰，因可能不易偵測、發現；尤其，再識別化行為倘在本國管轄範圍之外進行，更是如此[466]。即使能夠偵測與防止，科技總讓執法者落後進行再識別化者一步，在匿名化與再識別化的軍備競賽，後者常居上風[467]。學者[468]乃指出，禁止再識別化的禁令注定失靈，因難以偵測再識別化行為而不可能貫徹；再識別化常在不為人知的情形下進行。倘 Amazon 公司將客戶購物資料庫匿名化並傳輸予行銷業者；業者雖承諾不再識別化卻違反，他人可能知悉嗎？業者可以祕密進行再識別化而獲利，卻可能不被偵測到；雖可改採更嚴格處罰（如課予重罪或求償權）因應，仍障礙重重，任何因應措施的邊際效益終將為難以偵測的先天困境所侷限。然則，本文認為，數位資料（如個資[469]）或智財常屬無體性，各種侵害均面臨偵測的挑戰，個資立法的執行雖非完美但仍有一定嚇阻作

---

466 Phillips et al., *supra* note 447, at 533.

467 Angela L. Morrison, *A Research Revolution: Genetic Testing Consumers Become Research (and Privacy) Guinea Pigs*, 9 J. TELECOMM. & HIGH TECH. L. 573, 596 (2011).

468 Ohm, *supra* note 120, at 1758-59.

469 在數位時代，類似於其他無體資訊財產（如智慧財產），個資具非獨享、非互斥（nonrivalrous）特質，不會因某人的使用而即減損他人對於個資的重複使用，而且個資的使用或多人同時使用亦難以有效率地被排除；而這種個資的非排他性（non-excludable）特質乃因其虛擬性，無實體界線，其複製或分配所牽涉成本極低，排除他人使用的邊際成本常高於供他人使用的邊際成本。尤其，科技進步越來越多個資被產生、儲存且為他人控制，縱欲耗費資源追蹤每一筆個資下落，並藉以排除或追訴他人未經授權而使用個資之行為，並非易事，乃使得個資易於被過度使用、散佈，甚至濫用。翁清坤（2018），〈賦予當事人個人資料財產權地位之優勢與侷限：以美國法為中心〉，《臺大法學論叢》，47卷3期，頁952。

用；尤其，倘能配合憲法法庭 111 年憲判字第 13 號判決主文第二項之要求而如同其他諸多擁有個資法將各行各業均納入規範的國家（如歐盟會員國、加拿大、澳洲）設立獨立專責監督機關，將有助於各種違法濫用個資行為（如不當的再識別化行為）之偵測、發現，並有助於相關民刑事責任之追究。

另有質疑，實際運作上，再識別化的刑責將只針對某一部分民眾（如機構研究人員）而來，比起資料契約所承擔責任，刑責未必能產生更大嚇阻作用[470]。其實，臨床研究人員與研究參與民眾間常有醫病關係的適用，本應善盡注意義務（包括保守隱私在內），否則，將衍生後續責任[471]。

對於尚未可預知的未來研究，除事前廣泛取得資料主體同意之情形外，倘完全禁止再識別化，可能對於醫學研究形成寒蟬效應，也將對於民眾健康有不利影響。廣泛禁止再識別化將妨礙研究參與民眾近用對其健康有益的偶然發現（incidental findings），因再識別化與通報偶然發現將衍生民刑事責任而形成寒蟬效應，而使得研究者怯於利用可能產生偶然發現之研究技術。乃有主張，廣泛禁止再識別化不僅阻礙研究，對強化隱私保護亦有限[472]，故基於研究目的之再識別化應為禁止令之例外[473]。

---

470 Phillips et al., *supra* note 447, at 533.

471 Jonathan S. Miller, *How Did You Know That? Protecting Privacy Interests of Research Participants Via Certificates of Confidentiality*, 17 COLUM. SCI. & TECH. L. REV. 90, 111-12 (2015).

472 *Id*. at 113-114, 116.

473 Ahn, *supra* note 337, at 792.進而言之，如人體生物資料庫運作如產生「偶然發現」，國際實踐上，多數歐洲與美國資料庫目前以不提供此類發現之資訊予檢體的提供者為原則。因提供通知須將匿名檢體提供者再識別化，而須踐行告知後同意程序，成本遽增；但若不予通知易延遲治療，更損及民眾對此類研究之信任，進而影響參與研究意願及輿論支持。因此，一概課予通知義務、一概禁止通知，均非理想且合乎倫理的做法；在歐洲與北美的生醫倫理討論中，已逐漸形成「有限度通知義務」之共識。乃有主張除非檢體的提供者當時已於同意書明確拒絕受任何通知，否則可以「再接觸」方式徵詢是否有受偶然發現通知之意願。賈文宇（2018），〈人體生物資料庫通知基因研究「偶然發現」（incidental findings）之倫理及法律問題：兼論臺灣生物資料庫面臨之挑戰與建議〉，《政大法學評論》，153期，頁154-155、161、163-164。

另外，雖有時應強化資料濫用之制裁，但若由現行寬鬆規範跳躍至刑責，乃不合比例之舉，有違法律適用的一致性[474]。

（四）小結

去識別化資料仍可能再識別化，惟再識別化行為構成對於後來變成個資之不當蒐集，而牴觸本欲匿名化之目的。因此，對於濫用資料者課予民、刑事責任，以最小化再識別化風險，乃有主張應立法禁止惡意再識別化。惟有反對主張，縱對於再識別化行為入罪化，卻面對難以執行之挑戰，因可能不易偵測、發現；且實際運作上，其刑責將只針對某些人（如研究人員）而來，比起資料契約所承擔責任，刑責未必能產生更大嚇阻作用。然則，本文認為，數位資料均屬無實體性，各類侵害均面臨偵測難題，個資立法的執行雖非完美但仍有一定嚇阻作用。

陸、結 論

當前各產業與政府部門積極蒐集巨量資料，運用大數據技術進行各種研究分析，以導出創新性推論或發現，滿足社會各種需求。然而，資料增加流通運用同時，亦形成重大隱私風險。

在保護隱私與實現大數據之利益間，如何平衡，無疑一大挑戰。解決之道之一，即將個資去識別化、匿名化而不受個資法拘束，即可移作原始蒐集目的外利用或與第三人分享。原本隱私考量而被禁止利用之資料，因此開啟了新穎、第二次的利用，以供有益社會的各種用途。

惟大數據時代，常有眾多資料來源可供交叉比對，不論去識別化或匿名化資料均難以始終維持不可逆、不可還原的狀態，而不可避免均有被再識別化之風險，乃形成隱私等人格與經濟損害、表意自由的寒蟬效應。其實，去識別化若做得完善，在促進資料供各種公、私用途時，尚可保護隱私；若做

---

474 Phillips et al., *supra* note 447, at 533.

得不完善，則會侵害個人健康、尊嚴、聲譽或財務。因此，資料去識別化的進行應審慎為之。一般而言，去識別化之有效性，通常須視具體個案而定。去識別化通常非靠單一技術，而是不同方法、工具與演算法之集合而產生不同程度的有效性。惟去識別化方法越嚴謹，則資料效用將越低。

一些知名的再識別化事件，使得去識別化的有效性漸受質疑。但去識別化的擁護者則反駁，被再識別出來的比例實屬微小，去識別化仍屬有效機制。類似地，對於去識別化的有效性與再識別化風險，美國法院見解亦相當分歧。

針對上開爭論，本文建議可採取下列因應措施，以兼顧資料效用與隱私保護間的平衡，並降低再識別化風險：

（一）資料釋出不免須忍受被再識別化風險，立法例亦類似多採合理可能去識別化標準：因已被去識別化、匿名化資料仍可能與其他資料相結合，而有再識別化之可能；故較務實解決之道，應非在於完全排除再識別化風險，而是應著重於減緩風險至被再識別化可能性至極低之程度。由於資料釋出不免須忍受被再識別化風險，倘要求再識別化風險為零，將更難獲取可供研究資料而不利於公益。類似地，歐盟等國際立法普遍採用之「合理」識別化、去識別化標準，亦均未要求絕對地「完全排除被再識別化之風險」。對此，我國法雖未明定，法院見解分歧，但亦應採類似的合理標準為當。

（二）去識別化應兼採技術、行政與法律措施：釋出資料的再識別化風險越高，去識別化程度應越高。倘欲分享利用去識別化資料，則應結合數種措施、方法，以降低再識別化風險。因此，去識別化的進行應先按再識別化風險評估，而兼採符合比例之合理技術、行政與法律措施。

（三）課予民刑事責任而禁止不當再識別化：再識別化行為乃構成對於後來變成個資之不當蒐集，而牴觸本欲匿名化之目的。因此，對於濫用資料者課予民、刑事責任，以最小化被再識別化風險；縱面對執行之挑戰，仍有一定嚇阻作用。

參考文獻

一、中文部分

Maria Cristina Caldarola、Joachim Schrey（著），趙彥清、黃俊凱（譯）（2020），《大數據與法律實務指南》，元照。

Michele Wucker（著），許恬寧（譯）（2022），《找出生活中的灰犀牛：認識你的風險指紋，化危機為轉機》，天下文化。

Steven Pinker（著），陳岳辰（譯）（2022），《理性：人類最有效的認知工具，讓我們做出更好的選擇，採取更正確的行動》，商周。

Viktor Mayer-Schonberger、Kenneth Cukier（著），林俊宏（譯）（2013），《大數據》，天下文化。

王興娟（2018），〈行政院主計總處個資去識別化作業辦理情形與成效〉，《主計月刊》，753 期，頁 72-77，http://www.bas-association.org.tw/catalog/arts/010709072.pdf

江耀國、黃子宴（2019），〈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》，58 期，頁 1-62。

李寧修（2020），〈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〉，《臺大法學論叢》，49 卷 1 期，頁 1-50。https://doi.org/10.6199/NTULJ.202003_49(1).0001

吳全峰、許慧瑩（2018），〈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，頁 45-62。https://doi.org/10.3966/102559312018010272005

林裕嘉（2017），〈公務機關利用去識別化資料之風險評估及法律責任（上）〉，《司法周刊》，1852 期，頁 2-3。

--------（2017），〈公務機關利用去識別化資料之風險評估及法律責任（下）〉，《司法周刊》，1853 期，頁 2-3。

范姜真媺（2020），〈匿名加工資料制度之創設：因應大數據時代日本個人資料保護法之新進展〉，《東海大學法學研究》，59 期，頁 1-54。

財團法人電信技術中心（2017），《「通傳事業去識別化技術與相關技術規範研究」補助研究報告》，財團法人電信技術中心。

翁清坤（2013），〈告知後同意與消費者個人資料之保護〉，《臺北大學法學論叢》，87 期，頁 217-322。

--------（2018），〈賦予當事人個人資料財產權地位之優勢與侷限：以美國法為中心〉，《臺大法學論叢》，47 卷 3 期，頁 941-1051。https://doi.org/10.6199/NTULJ.201809_47(3).0001

--------（2020），〈大數據對於個人資料保護之挑戰與因應之道〉，《東吳法律學報》，31 卷 3 期，頁 79-159。

孫敏超（2021），〈美國加州修正加州消費者隱私法健康資料去識別化相關規定〉，《科技法律透析》，33 卷 1 期，頁 14-16。

張陳弘（2018），〈國家建置全民健康保險資料庫之資訊隱私保護爭議：評最高行政法院 106 年度判字第 54 號判決〉，《中原財經法學》，40 期，頁 185-257。

項靖、陳曉慧、楊東謀、羅晉（2015），《開放資料及其對政府治理與個人隱私影響之研究》，國家發展委員會。

賈文宇（2018），〈人體生物資料庫通知基因研究「偶然發現」（incidental findings）之倫理及法律問題：兼論臺灣生物資料庫面臨之挑戰與建議〉，《政大法學評論》，153 期，頁 145-191。https://doi.org/10.3966/102398202018060153003

樓一琳、何之行（2017），〈個人資料保護於雲端運算時代之法律爭議初探暨比較法分析：以健保資料為例〉，《臺大法學論叢》，46 卷 2 期，頁 339-422。https://doi.org/10.6199/NTULJ.2017.46.02.01

二、英文部分

Abbott, R. (2013). Big Data and Pharmacovigilance: Using Health Information Exchanges to Revolutionize Drug Safety. *Iowa Law Review, 99*, 225-292.

Achatz, C., & Hubbard, S. (2017). Us vs. Eu Guidelines for De-Identification, Anonymization, and Pseudonymization. *Journal of Internet Law, 20*(11), 1, 7-10.

Ahn, S. (2015). Whose Genome Is It Anyway?: Re-Identification and Privacy Protectionin Public and Participatory Genomics. *San Diego Law Review, 52*, 751-806.

Altman, M., Wood, A., O'Brien, D. R., Vadhan, S., & Gasser, U. (2015). Towards A Modern Approach to Privacy-Aware Government Data Releases. *Berkeley Technology Law Journal, 30*(3), 1967-2072. https://doi.org/10.15779/Z38FG17

Arbuckle, L., & El Emam, K. (2020). *Building an Anonymization Pipeline: Creating Safe Data*. O'Reilly Media.

Article 19. (2015, June). *Right to Online Anonymity*. https://www.article19.org/data/files/medialibrary/38006/Anonymity_and_encryption_report_A5_final-web.pdf

Ausloos, J. (2020). *The Right to Erasure in EU Data Protection Law*. Oxford University Press. https://doi.org/10.1093/oso/9780198847977.001.0001

Baron, J. B. (2012). Property as Control: The Case of Information. *Michigan Telecommunications and Technology Law Review, 18*(2), 367-418.

Benitez, K., & Malin, B. (2010). Evaluating Re-identification Risks with Respect to the HIPAA Privacy Rule. *Journal of the American Medical Informatics Association, 17*, 169-177.

Borgesius, F. Z., Gray, J., & van Eechoud, M. (2015). Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. *Berkeley Technology Law Journal, 30*(3), 2073-2131.

Brasher, E. A. (2018). Addressing the Failure of Anonymization: Guidance from the European Union's General Data Protection Regulation. *Columbia Business Law Review*, *2018*, 209-253. https://doi.org/10.7916/d8-zgve-y962

Brown, I., Wright, J., & Erdos, D. (2013). *Ethical Privacy Guidelines for Mobile Connectivity Measurements* (B. Zevenvergen, Ed.). University Of Oxford. https://doi.org/10.2139/ssrn.2356824

Brumfield, C., & Lee, J. J. (2020). The Risks and Rewards of Conducting A Census in the Digital Age. *Georgetown Law Technology Review*, *4*(2), 415-427.

Cate, F. H. (2010). Protecting Privacy in Health Research: the Limits of Individual Choice. *California Law Review*, *98*(6), 1765-1803.

Charkow, B. (2003). The Control over the De-Identification of Data. *Cardozo Arts & Entertainment Law Journal*, *21*(1), 195-228.

Cheung, A. S. Y. (2018). Moving Beyond Consent for Citizen Science in Big Data Health and Medical Research. *Northwestern Journal of Technology and Intellectual Property*, *16*(1), 15-40. https://doi.org/10.2139/ssrn.2943185

Chin, A., & Klinefelter, A. (2012). Differential Privacy As A Response to the Reidentification Threat: The Facebook Advertiser Case Study. *North Carolina Law Review*, *90*, 1417-1456.

Christovich, M. M. (2016). Why Should We Care What Fitbit Shares?: A Proposed Statutory Solution to Protect Sensitive Personal Fitness Information. *Hastings Communications and Entertainment Law Journal*, *38*(1), 91-116.

Cohen, J. E. (1996). A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace. *Connecticut Law Review*, *28*, 981-1039.

Contreras, J. L. (2016). Genetic Property. *Georgetown Law Journal*, *105*(1), 1-54.

Cunningham, M. (2017). Privacy Law That Does Not Protect Privacy, Forgetting the Right to Be Forgotten. *Buffalo Law Review*, *65*(3), 495-546.

Czarnowski, A. P., Kloc, K., Kunda, K., Gawronski, M., & Punda, P. (2019). CHAPTER 3 Security. In M. Gawronski (Ed.), *Guide to the GDPR* (pp. 185-231). Wolters Kluwer.

Dever, J. P., & Dever, C. J. A. (2017). A Democracy of Users. *Journal of Law & Cyber Warfare*, *6*(1), 8-50.

Deitch, J. (2020). Protecting Unprotected Data in Mhealth. *Northwestern Journal of Technology and Intellectual Property*, *18*(1), 107-128.

Drabiak, K. (2017). Caveat Emptor: How the Intersection of Big Data and Consumer Genomics Exponentially Increases Informational Privacy Risks. *Health Matrix: The Journal of Law-Medicine*, *27*(1), 143-228.

El Emam, K. (2013). *Guide to the De-identification of Personal Health Information*. CRC Press.

El Emam, K., & Arbuckle, L. (2013). *Anonymizing Health Data*. O'Reilly Media.

El Emam, K., Buckeridge, D., Tamblyn, R., Neisa, A., Jonker, E., & Verma, A. (2011). The re-identification risk of Canadians from longitudinal demographics. *BMC Medical Informatics and Decision Making*, *11*(46), 1-12.

European Union Agency for Fundamental Rights, European Court of Human Rights, European Data Protection Supervisor, & Council of Europe (2018). *Handbook on European Data Protection Law*. https://data.europa.eu/doi/10.2811/343461

Evans, B. J. (2013). Why the Common Rule Is Hard to Amend. *Indiana Health Law Review*, *10*(2), 365-414. https://doi.org/10.2139/ssrn.2183701

-------- (2016). Barbarians at the Gate: Consumer-Driven Health Data Commons and the Transformation of Citizen Science. *American Journal of Law and Medicine*, *42*(4), 651-685.

Federal Trade Commission (2012, March). *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers*.

https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf

Finch, K. (2016, April 25). *A Visual Guide to Practical Data De-identification*. Future of Privacy Forum. https://fpf.org/2016/04/25/a-visual-guide-to-practical-data-de-identification/

Finch, K., & Tene, O. (2014). Welcome to the Metropticon: Protecting Privacy in A Hyperconnected Town. *Fordham Urban Law Journal*, *41*, 1581-1615.

Finck, M., & Pallas, F. (2020). They who must not be identified: distinguishing personal from non-personal data under the GDPR. *International Data Privacy Law*, *10*(1), 11-36. https://doi.org/10.1093/idpl/ipz026

Froomkin, A. M. (1996). Flood Control on the Information Ocean: Living with Anonymity, Digital Cash, and Distributed Databases. *Journal of Law and Commerce*, *15*, 395-507.

Garfinkel, S. L. (2015). *De-Identification of Personal Information*. NIST. http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf

Gellert, R. (2020). *The Risk-Based Approach to Data Protection*. Oxford University Press. https://doi.org/10.1093/oso/9780198837718.001.0001

Gellman, R. (2010). The Deidentification Dilemma: A Legislative and Contractual Proposal. *Fordham Intellectual Property, Media & Entertainment Law Journal*, *21*(1), 33-61.

Gilligan, K. C. (2022). Protecting Consumers and Regulating Data: The Need for Comprehensive Federal Oversight of the Direct-to-Consumer Genetic Testing Industry. *Drexel Law Review*, *14*, 207-260.

Gitter, D. M. (2017). Informed Consent and Privacy of Non-Identified Bio-Specimens and Estimated Data: Lessons from Iceland and the United States in an Era of Computational Genomics. *Cardozo Law Review*, *38*(4), 1251-1299.

Groos, D., & van Veen, E.-B. (2020). Anonymised Data and the Rule of Law. *European Data Protection Law Review*, *6*(4), 498-508. https://doi.org/10.21552/edpl/2020/4/6

Hintze, M. (2016). *Viewing the GDPR Through a De-Identification Lens: A Tool for Clarification and Compliance*. https://fpf.org/wp-content/uploads/2016/11/M-Hintze-GDPR-Through-the-De-Identification-Lens-31-Oct-2016-002.pdf

-------- (2019). Science and Privacy: Data Protection Laws and Their Impact on Research. *Washington Journal of Law, Technology & Arts, 14*(2), 103-137.

Hirsch, D. D. (2014). The Glass House Effect: Big Data, the New Oil, and the Power of Analogy. *Maine Law Review*, *66*(2), 373-395.

Information and Privacy Commission New South Wales (2020, May). *A Guide to Privacy Impact Assessments*. https://www.ipc.nsw.gov.au/sites/default/files/2021-3/Guide_to_Privacy_Impact_Assessments_May_2020.pdf

Information and Privacy Commissioner of Ontario (2016, June). *De-identification Guidelines for Structured Data*. https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf

The Information Commissioner's Office (2012). *Anonymisation: Managing Data Protection Risk, Code of Practice*. https://ico.org.uk/media/1061/anonymisation-code.pdf

-------- (2017). *Big Data, Artificial Intelligence, Machine Learning and Data Protection*. https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

-------- (2021, October). *Chapter 2: How do we ensure anonymisation is effective? Draft anonymisation, pseudonymisation and privacy enhancing technologies*

guidance. https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf

Kish, L. J., & Topol, E. J. (2015). Unpatients: why patients should own their medical data. *Nature Biotechnology*, *33*, 921-924. https://doi.org/10.1038/nbt.3340

Klinefelter, A. (2011). When to Research Is to Reveal: The Growing Threat to Attorney and Client Confidentiality from Online Tracking. *Virginia Journal of Law and Technology*, *16*(1), 1-40.

Kloc, K., Gawronski, M., Dominiak, M., Sztaberek, M., Naklicka, P., & Punda, P. (2019). CHAPTER 1 Basic Compliance. In M. Gawronski (Ed.), *Guide to the GDPR* (pp. 3-115). Wolters Kluwer.

Klocke, J. L. (2008). Prescription Records for Sale: Privacy and Free Speech Issues Arising from the Sale of De-Identified Medical Data. *Idaho Law Review*, *44*, 511-536.

Lagos, Y. (2014). Taking the Personal Out of Data: Making Sense of De-Identification. *Indiana Law Review*, *48*(1), 187-203.

Lagos, Y., & Polonetsky, J. (2013). Public vs. Nonpublic Data: The Benefits of Administrative Controls. *Stanford Law Review Online*, *66*, 103-109.

Manheim, K., & Kaplan, L. (2019). Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law & Technology*, *21*, 106-188.

McCallister, E., Grance, T., & Scarfone, K. (2010). *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII): Recommendations of the National Institute of Standards and Technology*. NIST. https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-122.pdf

McGraw, D., & Leiter, A. (2012). A Policy and Technology Framework for Using Clinical Data to Improve Quality. *Houston Journal of Health Law & Policy*, *12*, 137-169.

-------- (2014). Risk-Based Regulation of Clinical Health Data Analytics. *Colorado Technology Law Journal*, *12*(2), 427-444.

Miller, J. S. (2015). How Did You Know That? Protecting Privacy Interests of Research Participants Via Certificates of Confidentiality. *The Columbia Science and Technology Law Review*, *17*(1), 90-119.

Morrison, A. L. (2011). A Research Revolution: Genetic Testing Consumers Become Research (and Privacy) Guinea Pigs. *Journal on Telecommunications and High Technology Law*, *9*, 573-605.

Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. In The IEEE Computer Society & The Institute of Electrical and Electronics Engineers, Inc. (Eds.), *2008 IEEE Symposium on Security and Privacy* (pp. 111-128). The IEEE Computer Society. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4531148

-------- (2010). Privacy and Security: Myths and Fallacies of "Personally Identifiable Information". *Communications of the ACM*, *53*(6), 24-26. https://doi.org/10.1145/1743546.1743558

National Committee on Vital and Health Statistics. (2017, February 23). *Recommendations on De-identification of Protected Health Information under HIPAA*. https://www.ncvhs.hhs.gov/wp-content/uploads/2013/12/2017-Ltr-Privacy-DeIdentification-Feb-23-Final-w-sig.pdf

National Data Guardian for Health and Care (2016). *Review of Data Security, Consent and Opt-outs.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF

NIST (2012, September). *Guide for Conducting Risk Assessments*. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf

Office of the Australian Information Commissioner (2017, May). *What Is Personal Information?*. https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/handling-personal-information?a=2832

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, *57*, 1701-1777.

-------- (2012, August 23). *Don't Build a Database of Ruin*. Harvard Business Review. https://hbr.org/2012/08/dont-build-a-database-of-ruin

-------- (2015). Sensitive Information. *Southern California Law Review*, *88*(5), 1125-1196.

Pan, S. B. (2016). Get to Know Me: Protecting Privacy and Autonomy Under Big Data's Penetrating Gaze. *Harvard Journal of Law & Technology*, *30*, 239-261.

Pavolotsky, J. (2013). Privacy in the Age of Big Data. *The Business Lawyer*, *69*(1), 217-225.

Peppet, S. R. (2014). Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent. *Texas Law Review*, *93*, 85-176.

Phillips, M., Dove, E. S., & Knoppers, B. M. (2017). Criminal Prohibition of Wrongful Re-identification: Legal Solution or Minefield for Big Data?. *Bioethical Inquiry*, *14*, 527-539. https://doi.org/10.1007/s11673-017-9806-9

Pike, E. R. (2016). Securing Sequences: Ensuring Adequate Protections for Genetic Samples in the Age of Big Data. *Cardozo Law Review*, *37*, 1977-2034.

Polonetsky, J., Tene, O., & Finch, K. (2016). Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification. *Santa Clara Law Review*, *56*(3), 593-629.

Porter, C. C. (2008). De-identified Data and Third Party Data Mining: The Risk of Re-Identification of Personal Information. *Shidler Journal of Law, Commerce & Technology*, *5*, 3-10.

Rostow, T. (2017). What Happens When an Acquaintance Buys Your Data?: A New Privacy Harm in the Age of Data Brokers. *Yale Journal on Regulation*, *34*, 667-707.

Rubinstein, I. S., & Hartzog, W. (2016). Anonymization and Risk. *Washington Law Review*, *91*, 703-760.

-------- (2017). The Anonymization Debate Should Be About Risk, Not Perfection. *Communications of the ACM*, *60*(5), 22-24.

Rubinstein, I. S. (2013). Big Data: The End of Privacy or a New Beginning?, *International Data Privacy Law*, *3*(2), 74-87. https://doi.org/10.1093/idpl/ips036

-------- (2016, November 8). *Brussels Privacy Symposium on Identifiability: Policy and Practical Solutions for Anonymisation and Pseudonymisation: Framing the Discussion.* https://fpf.org/wp-content/uploads/2016/11/Rubinstein_framing-paper.pdf

Russo, S. J. (2020). Is De-Identification of Personal Health Information in the Age of Artificial Intelligence A Reality or A Noble Myth?. *Journal of Health Care Compliance*, *22*, 55-59.

Schwartz, P. M., & Solove, D. J. (2011). The PII Problem: Privacy and A New Concept of Personally Identifiable Information. *New York University Law Review*, *86*, 1814-1894.

-------- (2014). Reconciling Personal Information in the United States and European Union, *California Law Review*, *102*(4), 877-916.

Segrist, P. (2015). How the Rise of Big Data and Predictive Analytics Are Changing the Attorney's Duty of Competence. *North Carolina Journal of Law & Technology*, *16*(3), 527-622.

Serwin, A. B. (2009). Privacy 3.0: the Principle of Proportionality. *University of Michigan Journal of Law Reform*, *42*(4), 869-930.

Smith, C. R. (2012). Somebody's Watching Me: Protecting Patient Privacy in Prescription Health Information. *Vermont Law Review*, *36*, 931-994.

Sokhansanj, B. A. (2012). Beyond Protecting Genetic Privacy: Understanding Genetic Discrimination Through Its Disparate Impact on Racial Minorities. *Columbia Journal of Race and Law*, *2*(2), 279-309. https://doi.org/10.7916/cjrl.v2i2.2276

Stalla-Bourdillon, S., & Knight, A. (2016). Anonymous Data v. Personal Data-A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data. *Wisconsin International Law Journal*, *34*(1), 284-322.

Sweeney, L. (2013). Matching Known Patients to Health Records in Washington State Data. https://privacytools.seas.harvard.edu/files/privacytools/files/1089-1.pdf

-------- (2002). k-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(5), 557-570. https://doi.org/10.1142/S0218488502001648

Sweeney, L., Yoo, J. S., Perovich, L., Boronow, K. E., Brown, P., & Brody, J. G. (2017). *Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study*. Technology Science. https://techscience.org/a/2017082801/

Swire, P., & Woo, J. (2018). Privacy and Cybersecurity Lessons at the Intersection of the Internet of Things and Police Body-Worn Cameras. *North Carolina Law Review*, *96*(5), 1475-1524. https://doi.org/10.2139/ssrn.3168089

Sylvester, D. J., & Lohr, S. (2005). Counting on Confidentiality: Legal and Statistical Approaches to Federal Privacy Law after the USA Patriot Act. *Wisconsin Law Review*, *2005*(4), 1033-1136.

Tene, O. (2013). Privacy Law's Midlife Crisis: A Critical Assessment of the Second Wave of Global Privacy Laws. *Ohio State Law Journal*, *74*(6), 1217-1261.

Tene, O., & Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property*, *11*(5), 239-273.

Teperdjian, R. (2020). The Puzzle of Squaring Blockchain with the General Data Protection Regulation. *Jurimetrics Journal*, *60*(3), 253-313.

The DHHS Office for Civil Rights. (2012). *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard

Tovino, S. A. (2004). The Use and Disclosure of Protected Health Information for Research Under the Hipaa Privacy Rule: Unrealized Patient Autonomy and Burdensome Government Regulation. *South Dakota Law Review*, *49*, 447-502.

Van Meter, B. T. (2020). Demanding Trust in the Private Genetic Data Market. *Cornell Law Review*, *105*(5), 1527-1560.

Verdi, J. (2012). Transcript: Sorrell v. Ims Health-Any Impact on Patient Privacy?. *Vermont Law Review*, *36*, 829-834.

Victoria State Government (2018). *De-identification Guidelines*. https://www.vic.gov.au/sites/default/files/2019-03/Victorian-Data-Sharing-Act-2017-De-identification-guidelines.pdf

Voigt, P., & von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer. https://doi.org/10.1007/978-3-319-57959-7

Voss, W. G., & Houser, K. A. (2019). Personal Data and the Gdpr: Providing A Competitive Advantage for U.S. Companies. *American Business Law Journal*, *56*(2), 287-344. https://doi.org/10.1111/ablj.12139

Waldo, J., Lin, H. S., & Millett, L. I. (Eds.). (2007). *Engaging Privacy and Information Technology in a Digital Age*. National Academies Press.

Warner, D. (2013). Safe De-Identification of Big Data Is Critical to Health Care Organizations Must Find A Way to Strike A Balance As They Work Through the Challenges and Concerns. *Journal of Health Care Compliance*, *15*, 63-72.

Westergren, A. (2016). The Data Liberation Movement: Regulation of Clinical Trial Data Sharing in the European Union and the United States. *Houston Journal of International Law*, *38*(3), 887-912.

Westin, A. F. (1967). *Privacy and Freedom*. Ig Publishing.

Woo, J. W. (2017). Smart Cities Pose Privacy Risks and Other Problems, but That Doesn't Mean We Shouldn't Build Them. *UMKC Law Review*, *85*(4), 953-972.

Wu, F. T. (2013). Defining Privacy and Utility in Data Sets, *University of Colorado Law Review*, *84*, 1117-1177. https://doi.org/10.2139/ssrn.2031808

Yakowitz, J. (2011). Tragedy of the Data Commons. *Harvard Journal of Law & Technology*, *25*(1), 1-68.

Yakowitz, J., & Barth-Jones, D. (2011). *The Illusory Privacy Problem in Sorrell v. IMS Health*. Technology Policy Institute. https://techpolicyinstitute.org/wp-content/uploads/2011/05/the-illusory-privacy-problem-i-2007545.pdf.

Zivanovic, N. N. (2015). Medical Information As A Hot Commodity: The Need for Stronger Protection of Patient Health Information. *Intellectual Property Law Bulletin*, *19*(2), 183-202.

The Deidentification of Personal Data and its Risk of Reidentification: A Legal Perspective

Ching-Kuen Ueng[*]

Abstract

The concept of "personal data" as the cornerstone for information privacy laws seems workable. Any data relating to an identified or identifiable natural person will trigger the mechanism of personal data protection.

The operation of big data is to derive or infer hidden value from the structured and unstructured raw data through novel reuse. However, the reuse of personal data will be likely beyond the scope of original collection purpose, in violation of the principle of purpose limitation. Furthermore, the ubiquitous use of personal data will lead to privacy risk. As a consequence, one of the solutions is to deidentify personal data in order to use for further purposes or share with third parties.

However, in the age of big data, as the deidentified or anonymized data may be combined with other datasets from various sources, it is not likely to absolutely ensure "a person cannot be identified from a dataset." The reidentification will cause damages to privacy, personality or property, and the chilling effect on freedom of expression.

As there were several famous reidentification cases in the past two decades, the effectiveness of deidentification or anonymization is gradually criticized. However, some scholars insist that the deidentification or anonymization is still effective in protecting privacy because the rate of reidentification is very small. Similarly, the U.S. courts are also divided in their effectiveness.

In facing the conflict between deidentification and reidentification, there could be some solutions. Firstly, the key point is to adopt a reasonable

[*] Associate Professor, School of Law, Fu Jen Catholic University.
E-mail: 071617@mail.fju.edu.tw

deidentification standard, thus reducing the risk of reidentification to a not important degree, rather absolutely ruling out its risk. Secondly, data controllers shall evaluate the risk of reidentification and thus adopt the technical, legal, and organizational safeguards subject to the principle of proportionality. Finally, statutes shall include civil and criminal liabilities in order to prohibit improper reidentification.

Keywords: personal data, privacy, big data, identifier, deidentification, anonymization, reidentification, freedom of expression, tolerant of risk, General Data Protection Regulation