司法院大法官審理會台字第 13769 號蔡季勳等申請案
意見資料

吳全峰（中研院法律所副研究員）

壹、 去識別化之爭議

在深入討論去識別化爭議前，有些名詞概念須先加以釐清，以利後續之討論：

（1） 無從直接或間接識別當事人：此為個資法之法律用語，為判斷資料是否屬個人資料並受個資法規範之條件。

（2） 去識別化資料：指經過處理（加密）後，資料科學上認可再識別之風險檻值降低到一定程度，而被認為達到「無從直接或間接識別當事人」條件之資料。但須注意以下兩點：

（A） 去識別化資料之概念是流動的：因為再識別技術之進步，可能使資料去識別化處理技術之再識別風險檻值產生變化，以前認為風險較低而為去識別化之資料，在新興技術下可能使再識別風險大幅提升，而不能再被認定是去識別化資料；且去識別化資料仍有被破解之可能，故即令該資料經過處理（加密），但若再識別風險仍高於檻值，則經過處理（加密）之資料仍不可被視為去識別化資料。因此，資料是否屬於去識別化資料，應該定期進行評估，且評估之方法與風險檻值亦應公開。

（B） 去識別化之判斷可以採統一判斷標準，亦可以採個案判斷標準。前者指所有資料之去識別化處理技術採統一標準，並認為採取該去識別化標準之資料，其再識別風險便已低於檻值；但採統一判斷標準所適用之資料集/庫，通常為封閉型資料集/庫，亦即該資料集/庫不會另外與其他資料集/庫串接，因為統一判斷標準對於再識別風險檻值之計算往往是以一定範圍內之資料數量與變項數量進行計算，若允許資料集/庫與其他資料集/庫進行串接，則資料數量與變項數量均將高於原本標準之設定，則再識別風險便可能產生誤差而不應適用。後者則是指針對每次申請使用之資料，依據其數量與變項進行再識別風險評估。

（3） 加密：主要指用來處理資料之統計或資料科學技術（如編碼、刪除變項、k-匿名（k-anonymity）、ℓ-多樣（ℓ -diversity）、差分隱私（differential privacy）、虛擬資料（synthetic data）等），但該技術是否能達到去識別化之標準則不一定；換言之，去識別化資料一定有經過加密處理，但經過加密處理之資料不一定為去識別化資料。舉例而言，假名化（pseudonymization，如將個人身份證字號以加密編碼取代）亦為加密技術之一種型態，但目前多已認為假名化資料仍可透

過與其他資料串接/對照以識別特定個人[1]（間接識別），故非屬去識別化資料[2]；另如，某加密技術針對 100 萬筆、50 個變項之資料庫足以達到去識別化之效果，但相同加密技術對於增加到 500 萬筆、100 個變項之資料庫，便可能無法達到去識別化效果。

就「無從直接或間接識別當事人」、「去識別化」、「加密」等基本名詞作出定義後，以下便針對健保資料庫之去識別化爭議進行討論。

（一）加密與去識別之概念不應混淆

就健保署交予衛福資料中心之資料如何加密及是否已去識別化之爭議，應從兩個層面加以觀察，首先為加密作為去識別化之工具，其內容與效果為何（議題一），其次則為加密之內容是否能達成顯著降低再識別風險之效果（議題二）。詳言之，若加密之效果有限，則自然無法達成去識別化之效果；但加密效果良好，也不代表再識別之風險便將顯著降低，加密之方式與加密之變項有許多種不同選擇，而再識別之途徑也有許多種，故目前健保署針對單一變項（如身分證字號）加密方式之妥適與否，並不代表再識別化之可能性便完全排除，而須經由整體評估（再識別風險評估）始能確認釋出資料之再識別之風險。舉例而言，一棟房子之前門防盜措施已經做到完善（議題一），僅能代表前門不易被盜賊侵入，但並不代表這棟房子之被盜風險就等於零（議題二），因為盜賊仍可能從窗戶、後門、氣窗等處侵入。

實務上，亦有不少案例可證明加密與再識別並不能畫上等號：

（1） 現為美國著名學者之 Latanya Sweeney 便展示多起完成加密/去識別的資料集，仍可與已知資料進行串連的案例[3]。Sweeney 最具代表性之 linkage attack 之案例，便是將經 HIPAA 安全港加密/去識別化後（僅留下 ZIP code、性別與生日）提供研究用途之資料集，與麻州州長公開資料與住院新聞訊息進行連結比對，即可再識別特定資料為州長之資料[4]。

（2） 在 Netflix Prize Data Study 之案例中，Netflix 釋出每個會員（以代碼表示，沒有姓名或社會安全號碼）對電影之評比紀錄（1-5分）與評比日期；就該資料庫而言，姓名與社會安全號碼之加密內容應為百分之百，因為已完全刪除相關資料。但在這個有限資料庫中，研究人員發現即令在完全沒有個人資料之情況下，若知道某個人對六部電影之準確評比，則有 84%可以再識別該個人之身份；如果研究人員知道該評比做出之大概時間（約兩週），則再識別之機率將高達 99%；且日期若能更為精確（約三日），則僅要知道兩部電影之評比即有 68%之機率可以再識別

[1] Emma Swahne, Data Protection by Design and Default, 64 (Master Thesis, University of Lapland, 2016).
[2] Article 23 of the General Data Protection Regulation (GDPR).
[3] Latanya Sweeney, Matching Known Patients to Health Records in Washington State Data, Harvard University. Data Privacy Lab. 1089-1, June 2013, https://arxiv.org/abs/1307.1370
[4] NIST, Simon L. Garfinkel, De-Identification of Personal Information (NISTIR 8053), P. 18, 2015, https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf

該個人[5]。換言之，加密之有效性並不一定能完全等同於再識別風險之降低（或隱私保障）。

（3） 近期則有澳洲學者於 2017 年之研究，將政府宣稱已完成加密/去識別化之健保資料加以測試，在不須將加密資料解密之前提下，仍可透過與個人相關且未加密之資料再識別 7 位澳洲公眾人物，並將其與加密/去識別化之健保資料加以串接[6]。澳洲健康部在事件發生之後馬上停止提供健康資料，並且再度評估開放敏感性特種個資使用的可行性。

因此，目前衛福部說明其針對健保署所傳送資料之機敏資料欄位代碼（如身分證字號、醫療院所代碼、申報日期、就醫日期等）進行第一次加密，並以流水碼對照方式進行第二次加密，於申請應用時再針對資料檔流水碼加密欄位一案號進行第三次加密，這些加密動作均係針對特定欄位重複進行加密，即令加密程序與硬體符合美國聯邦資訊處理標準 FIPS（Federal Information Processing Standards）140-2 Level 3 國際資訊安全標準，僅能表示這些特定機敏資料欄位之資訊安全保障已能達一定程度，而第三方可能不易重新還原這些欄位代碼（議題一）；但這些複雜之加密技術也僅止於還原這些機敏資料欄位之難度提升，但卻不能排除第三人可在不需還原這些欄位之前提下，透過比對其他資料之方式再識別特定個人（議題二）。健保署或稱其亦刪除敏感資料（如僅列出資料當事人之生年，並刪除日期），或可降低透過比對再識別之風險；但在目前科技下，相關技術仍可利用所謂「非敏感資料」進行再識別（如在 Netflix Prize Data Study 之案例中僅利用電影評比）。故歐盟「個人資料保護指令第二九條個人資料保護工作小組」(Article 29 Data Protection Working Party)於二〇一四年所發布之「第二一六號有關個人資料匿名技術意見書」（Opinion 05/2014 on Anonymisation Techniques，簡稱 WP 216）[7]便明確說明，資料控制者通常認為刪除或替換一個或多個識別因子便足以使資料集去識別化，但許多例子已表明情況並非如此，若準識別因子（quasi-identifiers）仍保留在資料集中，或資料屬性仍然能夠識別個人，那麼簡單地更改識別因子（如身分證字號）並不能避免再識別風險。

因此，衛福部或健保署以特定欄位（如身分證字號、醫療院所代碼、申報日期、就醫日期等）之加密技術是否完備，作為整體資料庫再識別風險之判斷基礎，實已將加密（議題一）與再識別風險（議題二）加以混淆；但在風險評估上，這兩項議題應分別判斷較為妥適。 尤其在再識別技術大幅進步之近代，去識別化技術早已跳脫傳統針對單一變項加密之模式，因爲對個人單一類別資料（如身分證字號）之重組（仍為目前健保署之主要加密技術），便被認為在目前之科技發展下可能已經無法因應隱私保障之挑戰[8]。近期發展

---

[5] See generally Arvind Narayanan & Vitaly Shmatikov, How to Break the Anonymity of the Netflix Prize Dataset, arVix, Oct. 16, 2006, at 1, http://arxiv.org/abs/cs/0610105v1 (v.1).

[6] Re-identification possible with Australian de-identified Medicare and PBS open data, https://www.zdnet.com/article/re-identification-possible-with-australian-de-identified-medicare-and-pbs-open-data/#:~:text=ZDNet%20Academy-,Re%2Didentification%20possible%20with%20Australian%20de%2Didentified%20Medicare%20and%20PBS,in%20an%20open%20medical%20dataset

[7] Article 29 Data Protection Working Party, Opinion 05/2014 on anonymization Techniques (hereinafter WP216) (2014), *available at* http://ec.europa.eu/justice/data-protection/article-29/documentation/ opinion-recommendation/files/2014/wp216_en.pdf (last visited Aug. 8, 2021).

[8] See e.g., Paul Ohm, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, 57 UCLA L. Rev. 1701, 1703-44 (2010).

之加密技術，如差分隱私（differential privacy）便是藉由數學計算公式增加非決定性雜訊（non-deterministic noise）方式以達到無從識別特定個人資料之目的[9]。但目前衛福部[10]或健保署[11]對去識別化之操作型定義仍是「指對特定欄位重新編碼加密，或重新模糊化，或予以增刪若干欄位，以無從辨識該特定個人」，此仍屬較為原始之去識別化定義，可能無法因應目前再識別化技術之進步。

（二） 再識別風險評估與風險檻值應透明化

另依目前之科技發展，我們必須承認並沒有任何技術可以百分之百排除再識別化風險；因此，僅能就特定資料庫經過哪些加密技術所可能達成之再識別風險進行評估（再識別風險評估），並以該風險之高低作為判斷是否符合個資法上「無從直接或間接識別特定當事人」之規範。但健保署在資料加密處理上，除前述將資料「經過加密作業」與「非屬個人資料」劃上等號，而未進一步檢視（或公開檢視數據）該加密資料之再識別風險是否已經低到「不具個人化屬性」之程度外，亦未與公民社會適當溝通再識別風險並建立公開透明之再識別風險評估機制，這種將「工具」（加密）等同於「目的」（無從識別當事人）、未適當建立風險溝通管道之方式，對於個人資料保護與隱私保障並不適當[12]。換言之，因為再識別之風險不可能為零，且為資料之可利用性亦不宜將再識別之風險設定為零，故如何透過與公民社會進行風險溝通、建立公開透明之去識別化機制、並設定可被社會接受之再識別風險檻值，便有其重要性，因其代表社會對於資料保護與資料利用間如何取得平衡之態度。

舉例而言，歐盟藥品管理局（European Medicines Agency，EMA）在 2018 年所發佈之歐洲藥品管理局人類用醫藥產品臨床資料政策實施外部指引（External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use）即說明，在進行個人資料去識別前必須先確認再識別風險檻值、並齊備再識別之風險評估；而設定可接受之再識別風險檻值，應包括對於現有未公開揭露資料集之評估、再識別風險在何種程度會被認為構成對資料當事人之隱私侵害、及攻擊者重新識別身份之動機與能力評估。而風險檻值一旦確定後，則可使資料集進行去識別化後之再識別風險可以實際概率加以測量，有助於在識別之風險評估：若概率高於風

[9] Simon L. Garfinkel, De-Identifying Government Datasets, 10 (NIST Special Publication 800-188), available at https://csrc.nist.gov/csrc/media/publications/sp/800-188/archive/2016-08-25/documents/sp800_188_draft.pdf (last visited May 3, 2020).

[10] 衛生福利部衛生福利資料科學中心資料申請應用規定及申請文件（110 年 6 月 1 日公告），https://www.mohw.gov.tw/dl-39218-b5baf813-79b5-4929-b930-281c09b67177.html

[11] 全民健康保險保險人資訊整合應用服務中心作業要點（110 年 4 月 20 日公告），https://www.nhi.gov.tw/DL.aspx?sitessn=292&u=LzAwMS9VcGxvYWQvMjkyL3JlbGZpbGUvMC81ODUzL%2bWFqOawkeWBpeW6t%2bS%2fnemaquS%2fnemaquS6uuizh%2bioiuaVtOWQiOaHieeUqOacjeWLmeS4reW%2fg%2bS9nOaIreimgem7nigxMTAuMDQuMjDlhazlkYopLnBkZg%3d%3d&n=5YWo5rCR5YGl5bq35L%2bd6Zqq5L%2bd6Zqq5Lq66LOH6KiK5pW05ZCI5oeJ55So5pyN5YuZ5Lit5b%2bD5L2c5qWt6KaB6bueKDExMC4wNC4yMOWFrOWRiikucGRm&ico%20=.pdf

[12] 吳全峰、許慧瑩，健保資料目的外利用之法律爭議--從去識別化作業工具談起，月旦法學雜誌，272 期，頁 45（2018 年）。

險檻值，則表示該資料集之利用需暫緩並進一步檢討加密/去識別之技術；而概率低於風險檻值，則表示資料僅具非常小的再識別風險（近於匿名）[13]。

但相對而言，衛福部或健保署對於健保資料庫並未公開其再識別風險評估之詳細機制、操作程序與結果，對於再識別風險檻值亦未與公民社會進行討論，導致公民社會對於攸關其健康之敏感性資料在利用過程中承受之再識別風險並無認識，更遑論提供公民社會足夠之資訊（充權）在隱私保障與資料利用之平衡（風險檻值之設定）進行實質討論。

至若衛福部所稱通過之 ISO27001 驗證證書亦僅係證明「獨立作業區域及雲端化網路資訊系統、語音資料統計管理作業系統運作與維護之安全管理」符合標準，但非再識別之風險評估。

（三） 去識別化之概念應細緻區分

進一步，去識別化之定義亦非僅以無法識別「特定」當事人為已足，依歐盟 Article 29 工作小組第 216 號有關匿名化技術意見書，資料經處理後是否達到匿名化（anonymization）狀態之三項判斷依據(風險)：

（1） 是否仍可能辨別特定個人（“single out” an individual）？
（2） 是否仍可連結至個人相關紀錄（link records “relating” to an individual）？
（3） 是否仍得從相關資訊推斷至個人（information can be “inferred” concerning an individual）？

若從歐盟之規範反觀健保署對於其所釋出資料之再識別風險判斷，衛福部或健保署似乎仍以「辨別特定個人」作為再識別風險之判斷標準，對於釋出資料是否可能「連結至個人相關紀錄」或「推斷個人」則似未有進一步說明。但對於去識別化過於單一之定義難稱妥適，因不同加密技術在前述不同再識別概念下之風險判斷並不相同（見表一），故針對不同資料庫在不同加密技術下產生之再識別風險，理論上應該要有較為細緻之說明與論述；但就目前健保署所釋出之公開資料中，未見到較為細緻之再識別風險評估報告，故無從得知目前健保署所採加密技術針對不同再識別風險之評估值為何。且就個資法第二條第一款所稱「直接或間接方式識別該個人」作為個人資料之判斷標準，亦難推導出僅有「辨別特定個人」方符合個資法之個人資料定義，若能從健保署所提供之資料庫中「連結至個人相關記錄」或「推斷個人」，則該資料仍應屬廣義上得「識別」個人之資料。

[13] EMA 基於 IOM 報告之建議，將風險檻值設定為 0.09。EMA, External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use, 49, 2018, https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-3.pdf

表一　去識別化技術之風險比較

	是否仍然存在被挑出的風險？	是否仍然存在關聯風險？	是否仍然存在推斷風險？
代碼化處理技術	是	是	是
添加雜訊	是	不太會	不太會
替換 (substitution)	是	是	不太會
聚合或 K-匿名	否	是	是
L-多樣性	否	是	不太會
差分隱私	不太會	不太會	不太會
散列函數/憑證化	是	是	不太會

（四）　不應以單一加密技術適用所有資料庫

最後一個問題是，能否以單一標準作為所有資料庫是否符合去識別化之判斷依據，答案則應該是否定的，因研究已指出是否可得再識別當事人並不適合單純以加密作業之種類做出判斷，而可能需要進一步依資料釋出之類型、數量與範圍做出判斷[14]，故並沒有一套放諸四海皆準的去識別化標準。詳言之，針對加密之硬體（如健保署所使用 FIPS 標準）、或針對單一變項（如屬連續變項之身分證字號）之加密技術可能比較容易發展出合適之統一標準；但針對資料數量、變項種類均不相同之不同資料庫，則很難有一套統一標準，因為資料數量多寡、變項種類之複雜程度，均可能影響加密技術之效能。

故如前文所述，採統一判斷標準所適用之資料庫，通常為封閉型資料庫，亦即該資料庫不會另外與其他資料庫串接。以美國健康保險可攜與責任法（Health Insurance Portability and Accountability Act，HIPAA）為例（此亦為台灣健保資料庫去識別化措施所主要參考之國外立法例），規範完成安全港（safe harbor）條款[15]或專家決定（expert determination）之去識別化（de-identified）程序[16]後之資料，已非屬受 HIPAA 保護之健康資料[17]，似乎是採單一標準。但必須注意者為：

（1）HIPAA 決定需移除與個人、個人之親屬、雇主、家屬相關之 18 項識別因子後，便可視為非可識別資料，是經過再識別風險評估後所決定；但考量 HIPAA 立法之主要規範對象限於「健康計畫、健康照護資料交換中心與健康照護提供者」（health plans, health care clearinghouse, and health care providers，covered entity），其資料庫之數量與變項內容與台灣單一保險人體制下之健保資料庫內

[14] Sophie Stalla-Bourdillon and 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, 284 (2016). See also C. Christine Porter, De-Identified Data and Third Party Data Mining: The Risk of Re-Identication of Personal Information, 5 Shidler J. L. Com. & Tech. 3, 3 (2008).
[15] 45 CFR §164.514(b)(2)(i).
[16] 45 CFR §164.514(b)(1).
[17] 45 CFR §164.502(d)(2).

容有不小差距（如全民健康保險保險人資訊整合應用服務中心之資料為「全國」而非單一或有限醫療院所之資料，而衛生福利資料科學中心可提供外界之資料檔清單中甚至包括 HIPAA 所規範資料庫中不會出現、且相對更為敏感之性侵害通報明細檔、兒少保護通報明細檔、家暴通報明細檔、身心障礙資料檔、中低收入戶資料檔等[18]，且均可申請全人口資料檔），故當初美國 HIPAA 對去除 18 項識別因子後之再識別風險評估，是否適合在沒有依據台灣健保資料庫之資料數量、變項種類重新進行再識別風險評估前，便平行移植至台灣健保資料庫，不無疑問。更何況 HIPPA 亦規定 cover Entities 與其商業夥伴尚須確認，對於刪除後之剩餘資訊，不論單獨使用或與其他資訊結合後，不具任何實際知悉特定個人身分[19]，亦即承認去除 18 項識別因子針對特定資料庫或資料利用方式，仍有較高之再識別風險，而需在去除 18 項識別因子後另作加密處理。至於 HIPAA 去識別化之標準近來飽受質疑，亦使美國開始思考重新檢討去識別化作業之模式[20]。

（2）在實務運作上，不同 covered entity 下之資料庫間不能互通（亦即同一份就醫資料集不可能同時存在 A 機構與 B 機構，除非經過當事人同意），因此亦不能串接，故利用單一加密標準作為控制再識別風險之方式，相對較為可行，因為其資料庫之數量、內容均相對單純；且雖然當事人可申請資料傳遞由 A 機構至 B 機構，但通常基於避免醫療糾紛之考量，B 機構往往會拒絕 A 機構之資料，故依 HIPAA 處理之資料庫通常並未與其他資料庫串接，加密作業上便相對單純。類似之案例在台灣健保資料庫之利用上，則可能不適用，如全民健康保險保險人資訊整合應用服務中心允許內部串接重大傷病檔，衛生福利資料科學中心所提供之資料亦允許與其他資料庫之資料進行串接比對[21]，使得以單一加密技術處理可能向外延伸串接之資料庫，並不適當。但健保署針對不同數量資料檔（兩百萬人 vs 全人口）、串接不同資料庫之資料檔，均採相同之加密技術，且未評估再識別風險是否有所差異，難稱允當。

因此，在個人資料保護上，應避免過度簡化且均一地將不同加密作業工具、不同資料庫內容賦予相同價值並做相同規範處理，以符合目前資訊科技快速進步下之個人資料保護概念。而台灣與美國 HIPAA 在個人資料去識別化機制上之其他差異，可進一步參考下表二。

---

[18] https://www.mohw.gov.tw/dl-17770-e13b366b-822c-4324-ba96-a9d76c053c36.html.

[19] 45 CFR § 164.514.

[20] Kathleen Benitez and Bradley Malin, *Evaluating Re-Identification Risks with Respect to the HIPAA Privacy Rule*, 17(2) J. AM. MED. INFORM. ASS. 169, 169-77 (2010). Latanya Sweeney et al., *Re-identification Risks in HIPAA Safe Harbor Data: A Study of Data FROM ONE ENVIRONMENTAL HEALTH STUDY*, 2017 TECH. SCI. 1, 1-75 (2017), available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6344041/pdf/nihms-988781.pdf (last visited Aug. 9. 2021).

[21] 張陳弘，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 9（2018 年）。

表二　美國 HIPPA 與台灣健保資料庫去識別化之差異

	美國	台灣
age	年，但 89 歲以上者僅能以一個類別出現（>= 90 yrs old）（45 CFR § 164.514(b)(2)(i)(C)）	僅為年，對高齡者無特別規定
Zip code	若居民人數超過 20,000 人，則提供 3 碼郵遞區號，若少於 20,000 人，則以 000 方式提供郵遞區號（45 CFR § 164.514(b)(2)(i)(B)(2)））	未考慮人數，從大於 200,000 人（新竹市東區）至 656 人（金門縣烏坵鄉）之分佈均有
概括條款	Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section（45 CFR § 164.514(b)(2)(i)(R)）	無
處理方式	需先去除 18 identifiers，故無法歸人（因為 name 等 identifiers 已刪除，而依 45 CFR § 164.514(b)(2)(ii)，covered entity 不能回溯刪除 identifiers 之資料集），亦無可串聯性；若要先歸人後再刪除 identifiers，則需當事人同意	身分證字號編碼後，同一人仍屬同一組號碼，故可歸人並進行串聯（不需當事人同意）

（五） 金鑰進行加密之資料集仍為可識別資料

依衛福部與健保署之說明，「健保署提供衛福部統計處之健保資料，係依衛福部之加密程式及金鑰進行加密處理」，並主張雖然「理論上承辦者[衛福部]可還原資料」，但因「依全民健康保險資料利用安全聲明第三點規定，未經健保署同意，健保資料不得解密還原成未加密資料使用」，故即「符合個人資料保護法施行細則第 17 條」無從識別當事人之規定。

但必需要說明者為，衛福部與健保署之說明係對去識別化概念之誤解。

（1） 以 WP 216 為例，其認為使用金鑰進行加密之資料集仍為可識別資料，因金鑰持有者可通過解密資料集輕鬆地再識別每個數據主體，故個人資料仍然被認為包含在資料集，而該資料集仍是可識別（假名）資料，而非去識別（匿名）資料。依 WP 216 之解釋，使用金鑰進行加密之資料集之所以不能被認為是去識別資料，是因為其再識別風險極高，而該風險本身便足以說明使用金鑰進行加密之資料集不符合「無從識別當事人」之條件；故使用金鑰進行加密之健保資料，便不能認為符合個人資料保護法施行細則第 17 條「無從識別當事人」之規定。

（2） 不論加密技術為何，資料控制者（data controller）均被要求負擔一定責任（accountability）確保加密資料（不論是否為可識別資料）之安全性。而全民健康保險資料利用安全聲明第三點所規定，衛福部未經健保署同意不得將健保

資料解密還原成未加密資料使用，便係對資料持有者責任之規範，並不應將對資料控制者責任之要求，轉化成為資料去識別化之判斷基礎，否則將使資料去識別化之客觀標準化定義出現不確定之狀況，並不利於個人資料保護。舉例而言，若主張「未經同意不得解密還原」可作為資料是否已去識別化之判斷標準，則資料控制者便不需完善加密技術或進行再識別風險評估，因只要在聲明中載明「未經同意不得解密還原」，則該資料便可逕自被認為非屬個人資料而不受個資法規範；但加密技術與再識別風險評估之欠缺，卻將導致個人資料保護之門檻被不適當地大幅降低，而與個資保護之本旨相違背。

（3） 因此，資料是否屬可識別資料，應從資料本身之屬性——亦即加密技術與該資料經加密後之再識別風險評估——加以判斷，而不宜與資料控制者之責任要求混為一談；故若加密之健保資料如衛福部所稱「承辦者仍可還原資料」，則該資料便屬可識別資料，而不符合個資法施行細則第 17 條「無從識別當事人」之規定。且以資料本身屬性（而非資料控制者責任）作為判斷資料是否去識別之基礎，不僅與國際規範、學理上去識別化概念相符，亦與個資法規範相符，因不論個資法第二條第一款獲個資法施行細則第十七條對個人資料之認定，均以「資料本身經過加密技術後是否無從辨識特定個人」作為判斷依據，而非以「不論資料本身經加密後是否仍得辨識特定個人，但資料控制者聲明不會辨識特定個人」作為判斷是否屬個人資料之依據。

貳、 學術研究是否屬公共利益之爭議

就健保資料庫之利用是否能以公共利益作為限制當事人資訊自主權之正當理由，此議題之核心不在於經過利益衡量後資訊自主權是否應退讓，因為隱私保護並非絕對權力，在追求適當公共利益之前提下，自有合理限制權利之空間；就法律明文規定、公務機關執行法定職務或非公務機關履行法定義務、協助公務機關執行法定職務或非公務機關履行法定義務之公共利益重要性等個資法所規範之例外規定，此處因限於篇幅不與深入分析。本文認為在個資不需當事人同意而得例外蒐集、處理與利用之公共利益判斷上，較值得爭執之問題在於，衛福部或健保署在處理個資法第六條第一項第四款、第十六條第五款、第十九條第四款、第二十條第五款所稱「學術研究機構基於公共利益為統計或學術研究」是否具有「必要」之判斷時，目前健保署或衛福部是否有設定適當並公開判斷標準與依據，且在作出決定後適度公開判斷之理由與依據。

以澳洲為例，其主要採「個案審議」加上「公共利益類型模組化」之方式對公共利益之正當性進行判斷。澳洲隱私法（Privacy Act）PART IV 授權澳洲資訊專員辦公室（Office of Australian Information Commissioner，OAIC）針對公務機關或非公務機關提出之申請進行公共利益審議，並判斷該公務機關或非公務機關之行為或作法所涉及之公共利益是否顯著重於（substantially outweigh）澳洲隱私法規所欲保護之利益[22]；OAIC 同時亦

[22] OAIC, About Public Interest Determination, https://www.oaic.gov.au/privacy/privacy-registers/public-interest-determinations-register/public-interest-determinations/

會公布公共利益審議指引 (Privacy Public Interest Determination Guide)，供公眾與其他具類似情境之公務機關或非公務機關申請時參考。而專員判斷具正當性之公共利益，又可分為大於 12 個月之公共利益決定（public interest determination，PID）或少於 12 個月之暫時性公共利益決定（temporary public interest determination），通過個案審議之公務機關或非公務機關可主張公共利益為正當化事由採取相關行為或作法，OAIC 亦將會把該個別申請案件內容與期限公告於公共利益註冊平台 （public interest register）[23]；且該公共利益審議亦受國會監督，國會得對公共利益決議之全部或一部提出 disallow 或 veto[24]。

英國則採「個案審議」加上「法律明文列舉」之方式。英國於 2018 年修訂之個人資料保護法主要依循歐盟 GDPR 之規定，而 GDPR 對公共利益之規定在適用時尚區分為一般公共利益（general public interest）與重大公共利益（substantial public interest）；雖然 GDPR 對於兩者之區別並未加以定義，但英國於 2018 年個人資料保護法 Schedule 1 Part 2[25] 列舉 23 項重大公共利益[26]類別（見下表三）用以補充說明 GDPR Article 9(2)(g)之規定[27]。資料管理者（data controller）欲主張重大公共利益作為敏感性特種個資蒐集、處理與利用之正當化事由[28]，其所主張之特定目的須符合 23 項重大公共利益類別之一，並須舉證其處理敏感性特種個人資料之必要性，在某些情況下還必須說明未能取得資料當事人同意之原因。另，個人資料保護法 Schedule 2 Part 4[29]要求資料管理者，主張公共利益作為處理敏感性個人資料正當化事由，必須具備適當政策文件，記錄資料管理者處理敏感性特種個資之重大公共利益類別、遵循個人資料保護原則的程序、保留和刪除個人資料的政策，及保留特定個資的期限等，作為有責之證明文件[30]。惟是否能夠通過重大公共利益的測試，還是要以個案判斷[31]。

---

[23] OAIC, Public Interest Determinations Register, https://www.oaic.gov.au/privacy/privacy-registers/public-interest-determinations-register/?start=0&year=2019

[24] OAIC, Privacy Public Interest Determination Guide, https://www.oaic.gov.au/privacy/guidance-and-advice/privacy-public-interest-determination-guide/

[25] UK Personal Data Protection Act 2018, Paragraph 6 to 28 of Schedule 1.

[26] ICO Guidance, Lawful Basis for Processing Special Category Personal Data, 15, 2019, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data/

[27] ICO, Lawful Basis for Processing Special Category Personal Data, 35-35, 2019, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data/

[28] GDPR Article 9(2)(g), processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

[29] ICO, Personal Data Protection Act, Schedule 2 Part 4.

[30] ICO, Guidance, Lawful Basis for Processing Special Category Personal Data, 36-37, 2019, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data/

[31] ICO, Guide to General Data Protection Regulation, 84, 2021, https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr-1-1.pdf

表三　英國年個人資料保護法之重大公共利益類別

Conditions	Show substantial public interest	Justify why no consent	Appropriate policy document
Statutory and government purposes	Y	N	Y
Administration of justice and parliamentary purposes	N	N	Y
Equality of opportunity or treatment	N	N	Y
Racial and ethnic diversity at senior levels	N	Y	Y
Preventing or detecting unlawful acts	Y	Y	Y/N*
Protecting the public	Y	Y	Y
Regulatory requirements	Y	Y	Y
Journalism, academia, art and literature	Y	N	N
Preventing fraud	N	N	Y
Suspicion of terrorist financing or money laundering	N	N	Y
Support for individuals with a particular disability or medical condition	Y	Y	Y
Counselling	Y	Y	Y
Safeguarding of children and individuals at risk	Y	Y	Y
Safeguarding of economic well-being of certain individuals	Y	Y	Y
Insurance	Y	Y	Y
Occupational pensions	N	Y	Y
Political parties	N	N	Y
Elected representatives responding to requests	N	Y	Y
Disclosure to elected representatives	N	Y	Y
Informing elected representatives about prisoners	N	N	Y
Publication of legal judgement	N	N	Y
Anti-doping in sport	N	N	Y/N*
Standards of behaviour in sport	Y	Y	Y

若從澳洲與英國之案例觀察，個案判斷為公共利益判斷之重要方式，因為公共利益具有複雜之多樣內涵；而台灣健保資料庫之釋出基本上亦採個案判斷之方式，誠如法務部104年7月2日法律字第10403508020號函所稱，公共利益「係指為社會不特定之多數人可以分享之利益而言」，其中學術研究指「針對有系統而較專門之學問中的特定主題，作深入且有系統的探討或研查 ，以發現事實，形成理論並付諸應用，但因屬「抽象之法律概念，故尚難遽定其範圍及認定標準，仍宜依具體個案分別認定之」。全民健康保險保險人資訊整合應用服務中心之資料，依「全民健康保險保險人資訊整合應用服務申請案件審核作業原則」規定，由健保署加以審核（必要時得請外部專家辦理複審）；衛生福利資料科學中心之資料，依「衛生福利部衛生福利資料申請案件審核作業原則」，由衛福部統計處（必要時納入外部專家）進行審查。

但相較於相較於澳洲與英國案例，台灣健保資料庫之利用在公共利益之判斷上，仍有以下幾點可供思考：

（1）可歸因性（relevance）：不論是澳洲 Privacy Public Interest Determination Guide 或是英國個人資料保護法 Schedule 1 Part 2，對於公共利益之判斷均有提供基本之判斷方向或準則，英國甚至未直接將學術研究列為重大公共利益（學術研究構成重大公共利益尚須符合特定要件，如與非法或不誠實行為有關）[32]；但在台灣，不論「全民健康保險保險人資訊整合應用服務申請案件審核作業原則」或「衛生福利部衛生福利資料申請案件審核作業原則」，均未對得限制資訊自主權之公共利益內容（尤其是學術研究內容是否具必要性）應如何判斷，提供基本之重要評估方向或準則（可為原則性規定或基本操作型定義即可），而此規範上之真空便造成如何判斷申請健保資料庫利用之學術研究具有必要性、是否存在其他適合替代方案，缺乏評估之判準參數。

在此情況下，不難發現衛福部或健保署所核准之計畫雖然包含學術單位、醫療院所及衛生福利相關學、協會等，但係就其類型則包括期刊、碩博士論文、研討會論文/海報、研究報告等，而研究內容雖然不乏具重要學術價值之研究議題（如健保署回覆意見中所提），但亦包括如「提升國人氣候變遷健康適能與調適策略研究」、「新聞情緒指標與醫療利用率之關聯性」、「名人效應如何影響民眾的醫療利用」等較不具急迫性之研究主題（參考衛生福利資料科學中心研究成果登錄系統），而不同案件與不同研究主題之公共利益必要性否相同，便有討論空間；但因衛福部與健保署在此均未有適當公開之指引或說明，從而導致這些不同類型、不同研究主題學術研究之「必要性」均被判斷為同一，可能並非允當。

另須說明者為，不可否認所有研究類型與主題均有其學術價值與重要性，但此處之核心問題在於是否「所有」研究主題均重要到足以實質限制（亦即澳洲在PID所稱之「substantially outweigh」）個人自主選擇其健保資料不被利用之資訊自主權。若以衛福部與健保署針對民眾得要求退出健保資料庫之說明觀察，其

[32] paragraphs 13 of Schedule 1 of the DPA 2018.

理由之一便是允許民眾退出可能提升研究門檻；但對於無立即風險、對國民健康影響有限、對政策執行無實質影響之研究，是否需要藉由限制民眾之資訊自主權以降低這些相對公共利益較低之學術研究之研究門檻，可能會面臨較大之挑戰。

（2）公開性（publicity/transparency）：澳洲與英國均會將公共利益審議之申請文件、諮詢與審議過程、結論，經過適當程序（如遮蔽部分資訊）後公開，以其社會能理解或檢視其公共利益判斷之理由，並判斷其與前述判斷公共利益之基本原則是否具連貫性、整體性與一致性。但在台灣，不論是「全民健康保險保險人資訊整合應用服務申請案件審核作業原則」或「衛生福利部衛生福利資料申請案件審核作業原則」均未規定審查會議記錄之公開；而就目前之公開系統（以衛生福利資料科學中心研究成果登錄系統[33]為例）可查詢者僅包括最新成果與最新計畫，且除計畫名稱、申請人、服務機關、委託單位、計畫摘要及申請欄位檔案外，並無該通過計畫是否具有足夠限制資訊自主權之公共利益之具體判斷理由，有些計畫甚至缺少申請欄位檔案之資料。

參、 得否允許退出健保資料庫之爭議

在此需先釐清一個概念，就健保署因公務機關執行法定職務必要範圍內（個資法第六條第一項但書第二款）所蒐集之個人資料所成立之資料庫，應無選擇退出（opt out）之選項，因健保署在個人強制加入健保之過程中，對其法定職務將持續執行並有蒐集、處理與利用個人資料之必要性。故此處問題之癥結，並非「個人得否選擇退出健保資料庫」，而是「個人是否能選擇其健保資料庫之個人資料利用，僅供健保署執行法定職務使用、而不同意給予學術研究做二次利用」；換言之，健保資料庫之完整性仍然存在，僅是在利用目的上做出區隔。

就不同類型學術研究之公共利益是否均可被認為重要到超過對個人資訊自主權之保障，本文已於前文（第貳節）做過說明，並認為需要有較細緻之討論與區分而不宜一蓋視為同一；此處需要進一步說明者為，允許個人選擇不同意健保資料庫之檔案予學術研究做二次利用，是否具有正當性。就目前衛福部與健保署對此議題之說明，大致可分為以下幾點：

（1）資料完整性：衛福部與健保署均主張允許民眾選擇其資料不予學術研究進行二次利用，可能影響資料完整性與取樣偏誤，並帶來不確定之風險與提升研究門檻。但就衛福部與健保署之主張，可能需要面臨以下之挑戰：

[33] 參考 https://rais.mohw.gov.tw/#/pages/home。

（A） 目前申請健保資料庫之學術研究案件可分為兩類，兩百萬人抽樣檔與全人口資料檔，則既然兩百萬人檔已能滿足資料完整性與研究門檻之要求，則允許部分民眾選擇退出不參加學術研究，應亦不至於影響資料完整性與研究門檻。

（B） 兩百萬人檔之選擇係依據母群體（全人口）之特徵以統計方式選出（按比例隨機）樣本，以避免兩百萬人檔與全人口檔產生抽樣誤差；同理，在衛福部與健保署仍因法定職務需要保有完整之全人口健保資料庫時，即令部分民眾選擇不同意學術研究對其資料之二次利用，衛福部與健保署仍有統計方法自母群體（全人口健保資料庫）抽樣出不含退出民眾、且無抽樣誤差之兩百萬人資料檔。

（C） 誠如衛福部之回覆意見，允許個人不同意學術研究二次利用之影響，較為嚴重者應為罕見疾病、特定年齡層等小範圍之資料，因母群體相對較小，故較可能產生收樣偏誤；但須注意者為，這類小範圍資料之再識別風險本就偏高，因此如美國 HIPAA 規範下針對這類小範圍族群進行研究時，往往會有特別規範（如透過專家核決以確認經處理後資料僅有非常微小之再識別風險）[34]或重大公共利益需要。因此，就罕見疾病等特殊族群之資訊自主權之限制，在再識別風險偏高之前提下，不宜僅以取樣偏誤作為正當化基礎，而需有更嚴格之規範，如再識別風險之評估與資訊安全機制應更為嚴格、無其他侵害更小之取樣方式等。

（D） 而對學術研究若選擇使用全人口資料檔，是否可能受到民眾選擇退出學術研究二次利用資料之影響？答案應該是肯定的，但因為全人口資料檔所包含之資料內容更為完整，同樣可能提高再識別之風險，故同樣不宜僅以取樣偏誤作為限制民眾資訊自主權之正當化基礎，而需有更嚴格之規範，如再識別風險之評估與資訊安全機制應更為嚴格、無其他侵害更小之取樣方式等。其實在「衛生福利部衛生福利資料申請案件審核作業原則」中已可見類似之規範方向，其就兩百萬人檔與全人口檔申請案之審查程序嚴格程度並不相同（參考審核作業原則第三點）；因此，在申請使用全人口資料檔時，除非有明確且重大之公共利益，否則仍應保障民眾自由選擇是否同意學術研究對其健保資料二次利用之權利，且此方向應與目前衛福部區別兩百萬人抽樣檔與全人口資料檔之審查機制一致。

（2） 搭便車效應：健保署主張允許民眾選擇不參與學術研究將使其不付成本而坐享他人之力，故反對退出機制；但此主張卻與生命倫理之基本精神相違背。蓋依自主原則與知情同意原則，在面對研究可能有之風險（在健保資料庫案中即為再識別之風險），民眾應有自主選擇是否加入研究之權利，甚至有「拒絕參與或隨時撤回同意而不受報復之權利」（World Medical Association Declaration of

[34] CFR §§164.514 (c) - (h).

Helsinki第26點）；而國家或研究者均不宜以研究具有公益性，便強制民眾加入研究，因為該強制行為將構成「否定具自主能力個人之深思熟慮想法並剝奪其依其想法為行為之自由」，而違反 Belmont Report 所強調對個人之尊重，亦違反 Declaration of Helsinki 所強調「具有知情同意能力者在尚未全然同意之前，不得被納入研究」之知情同意原則。更進一步，若主張因搭便車效應而強制所有人均應參加該研究，亦可能違反知情同意之「自願」要素（Belmont Report）與違反「不能因病人拒絕參與或中途退出研究而受負面影響」之原則（Declaration of Helsinki 第31點）。

且實務上，不論臨床試驗、人體試驗或人體研究，均採自願參與之模式，亦未見因發展之藥物具有高度公益性，便允許強制民眾參與相關試驗，甚至指責未參與試驗者為「不付成本而坐享他人之利」；且學術研究之相關研究所得之成果，仍是全民共享之公共利益。

最後，針對雖然衛福部與健保署主張不允許民眾選擇退出學術研究對其健保資料之二次利用，將有助於提升醫療衛生發展，本文希望能提供一個比較不一樣的想法：

（1） 在當事人同意之情況下——可能以推定同意（presumed consent）與選擇退出（opt-out）之機制滿足當事人同意之要件——將可提供更廣泛之資料可利用性；雖然開放細節可能仍需要再行細部研議，但根據同意內容，資料之利用仍將較目前之機制享有更大之開放空間與多元可能性。因為在目前衛福部與健保署之制度設計中，學術研究對健保資料之二次利用係在未經當事人同意之前提下進行，依個資法之規定便必須進行資料去識別化，而去識別化資料之可利用性反而將會大幅降低（因為去識別化作業將導致一些資料被刪除或模糊化），反不利於達成衛福部與健保署所主張提升醫療衛生發展、促進學術研究之目的。

（2） 且為確保去識別化之機制順利運作並降低再識別之風險，衛福部與健保署必須採取相對嚴格之標準，如要求使用者僅能在實體隔離獨立作業區進行分析；但衛福部與健保署之硬體與軟體設備並不一定能滿足研究者之需要，反而限制或妨礙學術研究之可能性與進步。但若透過充分告知當事人資料釋出之再識別風險並取得其同意，則資料便可不用限制於實體隔離獨立作業區進行，反而更能夠發展合適之研究方法與分析技術。因此，對於許多研究者而言，若能藉由建立當事人同意機制以達到放寬資料使用限制之目的，對於學術研究發展可能反而是助力而非阻力。

肆、 健保資料庫使用者所能看見內容之爭議

（一） 仍可看見個人之單筆資料串

雖然衛福部與健保署均強調使用者僅能攜出經審核通過之整體統計結果，而無法攜出個別資料；但在實際操作上，使用者在實體隔離獨立作業區所能看到之資料仍是個別之單筆資料串（見附件一），僅身分證字號經過加密處理，並針對日期等必像做模糊化處理。而這些個別資料尚能進一步進行內部串接（因同一組身分證字號經加密處理後為同一串代碼，故可以該代碼為基礎進行資料串接）或歸人（指整理同一人在健保資料庫中數年間或數十年間之所有住院、門診、慢性處方箋等健保資料）；也因為除身分證字號外之其他資料均為真實，因此雖然身分證字號之加密技術嚴謹而難以破解，使用者仍有可能透過其他真實、個別資料之比對，使得再識別風險大幅提高（詳前文第壹節之說明）。

而在附件一之資料中不難發現，可透過不同變項之交互選擇，而使得篩檢出之人數逐漸降低，或許無法達到「辨別特定個人」之結果，但仍可能有「連結至個人相關紀錄」或「推斷個人」之再識別風險。舉例而言，依衛福部所公布之資料庫使用手冊，在「全民健保處方及治療明細檔_門急診-西醫、中醫及牙醫（HEALTH-01: H_NHI_OPDTE）」，在全屬別（HOS）變項中可以限定條件為「公立」「榮民醫院」，在醫療機構所在地（CITY）變項中可以限定醫療機構所在鄉鎮之條件（代碼可參照醫療機構現況檔（（H_OST_RESMF)，如設定為台北市北投區），則便能將範圍侷限在台北市立榮民總醫院之病人；藉由類似之操作，便能逐步縮小個案數量。雖然類似之操作可透過實體隔離獨立作業區對使用者操作行為之監督加以避免，但不可否認再識別之風險仍然可能存在，故再識別風險評估仍有其重要性。

（二） 與外部資料庫轉連後，可看見之個人資料串內容將增加

進一步，如前文第壹節在去識別資料之討論中，曾提及若允許資料集/庫與其他資料集/庫進行串接，則資料數量與變項數量均將高於原本標準之設定，則再識別風險便可能大幅度提高；故以美國 HIPAA 而言，少見不同資料庫串連之情形，以避免再識別風險高於可允許之範圍。

但在健保資料庫在學術研究之二次利用上，卻允許資料庫之外部串接，不僅健保資料庫使用者所得看到之個別單筆資料串之內容（亦即可以看到之變項數量）將較原本健保資料之內容更為增加，再識別之風險亦可能提高。詳言之，依衛生福利資料科學中心作業須知第四（八）點之規定，若符合衛署醫字第 1010265083 號函公告「得免取得研究對象同意之人體研究案件範圍」之第三款及第四款規定免取得受試者同意書時[35]，則健保資料庫使用者可攜入特定分析對象檔案與健保資料庫進行串聯；另從衛福部所通過之研究計畫中，亦不難發現健保資料庫與其他部會資料庫進行串連之情形，如 H107115（國衛院委託）便攜入勞動部農保資料、勞保資料與健保資料庫進行串連，H108168（中研院計畫）

[35] 主要指「研究屬最低風險，對研究對象之可能風險不超過未參與研究者，且免除事先取得同意並不影響研究對象之權益」與「研究屬最低風險，對研究對象之可能風險不超過未參與研究者，不免除事先取得研究對象同意則無法進行，且不影響研究對象之權益」兩種情形。

便攜入主計處人口普查調查資料檔與健保資料庫進行串連，研究案 H109257（交通部運輸研究 所 委託）便攜入內政部警政署交通事故資料庫與健保資料庫進行串連。而健保資料庫與外部資料庫經過串連後，則資料庫之內容將會增加，再識別之風險亦將增加。

附件一

由於目前衛福部資料科學中心之模擬資料庫需申請方得使用，故目前僅能使用國衛院早期釋出之練習用虛擬資料檔（97/08/29 上線，但目前已無法於網路上取得），並搭配新舊版譯碼簿（資科中心 codebook 與國衛院 codebook）進行說明；但因譯碼簿之英文欄位名稱並未有太大變化，故國衛院之虛擬資料檔應仍有參考價值，而兩者差距部分會在出現時做特別說明。

另，國衛院之虛擬資料檔目前連結內的資料多數為 SAS 檔，沒有軟體應該無法查看，故此處資料已使用轉檔軟體將資料（以 2001 年資料為例）轉成 EXCEL 檔，方便點選查看。

1. 進入加值中心後，研究者可以看到的資料為每一個個案之相關資料，並非處理完之統計資料，且除身分證字號做過加密處理外，其餘資料均為真實資料。
   且同一組身分證字號，加密號仍為同一，故可進行歸人，並在資料庫中看到同一個人的每一筆承保、門診、住院、醫生所開立治療方式（醫令）等資料。
   先以國衛院承保資料檔為例（資科中心有相同之資料檔）：

保險對象身分證字號，經過加密處理 現行編碼為ID

被保險人身分證字號，經過加密處理，與被保險人身分證字號不同者為眷屬；現行編碼為ID1

保險對象出生年月日 現行編碼為ID_BIRTH_Y ，僅至出生年，無月日

保險對象性別，現行編碼為ID_S

投保單位區域代碼，現行編碼為ID1_CITY，代碼參考醫療機構現況檔，最小可至新鎮市區

每一行表示單一人之單次承保紀錄

此兩行表示保險對象與被保險人為同一人(編碼相同)，為此人之兩次承保紀錄

此五筆表示同一保險對象，一共有三位被保險人(編碼相同有三筆)，為保險對象之五次承保紀錄(承保時依附之被保險人不同)

ID	INS_ID	INS_ID_TYPE	INS_AMT	ID_BIRTHDAY	ID_SEX	INS_RELATION	UNIT_INS_TYPE	AREA_NO_I	ID_IN_TYPE	ID_IN_DATE	ID_OUT_TYPE	ID_OUT_DATE
0006a51471086bf0c5709f53e28e64b9	0006a51471086bf0c5709f53e28e64b9	2	19200	19550215	M		21	3408	1	19950301		
012e61103b1db1b6e92ecf423de53aa5	7a9a073290b977a48ea6f171ea9492ed		0	19820921	M	3	21	109	1	19950301		
014fd9cb0edfc5b57255ee4a9e39660b	014fd9cb0edfc5b57255ee4a9e39660b	2	21900	19740327	M		11B	118	1	19950301		
015c3366e165cec2e8150c3941b28a2f	015c3366e165cec2e8150c3941b28a2f	2	30300	19700101	M		12H	3205	1	19971127	1	20001018
01adaa2d0a09b8e6781ebc0bcbc23d4d	17c99b4cc5b43cdf503397155bc7272c		0	19810520	F	3	62	4314	1	19950529	2	19990127
01adaa2d0a09b8e6781ebc0bcbc23d4d	17c99b4cc5b43cdf503397155bc7272c		0	19810520	F	3	31Q	4314	2	19990127		
025ce9868ffe69b3ea57018f64401c77	4c7ede3ce1c94f1bfda0f3062958a936		0	19521115	F	2	12H	101	1	20000323		
025ce9868ffe69b3ea57018f64401c77	025ce9868ffe69b3ea57018f64401c77	2	14400	19680724	F		12H	2106	1	19950301	2	19950718
025ce9868ffe69b3ea57018f64401c77	f5795db90a79058fc43650e778761906		0	19680724	F	1	62	2106	2	19950801	2	19990426
025ce9868ffe69b3ea57018f64401c77	025ce9868ffe69b3ea57018f64401c77	2	1007	19680724	F		62	4313	1	19990427	2	19990707
025ce9868ffe69b3ea57018f64401c77	025ce9868ffe69b3ea57018f64401c77	2	19200	19680724	F		31Q	4313	2	19990729		
027338268da87ff41e7a17f65854edab	5d79d9543f3dd8ce0e1af951289d2286		0	19801111	M	3	12H	3715	1	19950301	2	19970115
027338268da87ff41e7a17f65854edab	027338268da87ff41e7a17f65854edab	2	910	19801111	M		62	1706	3	19970101	4	19970401
027338268da87ff41e7a17f65854edab	5d79d9543f3dd8ce0e1af951289d2286		0	19801111	M	3	31Q	3705	5	19970429		

2. 再以國衛院住院醫療費用清單明細檔（DD）節錄資料為例，此檔案型態類似現制衛福部資料中心之全民健保處方及治療明細檔–西醫住院檔（HEALTH-02：H_NHI_IPDTE）：

保險對象身分證字號，經過加密處理，現行編碼同樣為ID
經整理後，此12筆資料為同一保險對象，而此資料則顯示該保險對象於一段時間內共住院12次，而每一行均表示每一次住院之相關資料，內容包括但不限於以下項目：

- 性別
- 醫師身分證字號(加密)及性別
- 就醫科別
- 入院日期及入院、出院年齡
- 住院天數(急性或慢性病床)
- 診斷代碼(ICD9-CM)
- 手術處置(ICD9-PCS)
- 就醫醫療機構代碼(加密)與權屬別及所在縣市鄉鎮區
- 給付類別(如包括職業傷害、職業病、普通傷害、普通疾病、剖腹生產、呼吸照護、安寧照護、高雄氣爆事件、新加坡航空空難、氣喘醫療給付改善方案等)
- 是否免部分負擔(包括免部分負擔之理由為重大傷病、分娩、低收入戶、榮民或遺眷家戶代表、多氯聯苯中毒油症患者)
- 依附就醫新生兒日齡
- 是否為汽車交通事故
- 案件分類：是否為AIDS案件、結核病痰抹片楊姓<=14天案件、結核病MDR及抗結核藥物副作用住院<=60天案件、低收入戶住院膳食費案件等

"全民健保處方及治療明細檔–西醫住院檔"無此欄位，但有年齡欄位(編碼為IN_AGE, OUT_AGE)
操作上仍可利用前欄ID與前點所述之"承保檔"串結，即可得到保險對象出生年，而現行編碼為ID_BIRTH_Y

給付類別(現行編碼同)，包括職業傷害、職業病、普通傷害、普通疾病、剖腹生產、呼吸照護、安寧照護、高雄氣爆事件、新加坡航空空難、氣喘醫療給付改善方案等

就醫科別(現行編碼同)

出院日期(現行編碼無)

入院日期(加密處理)

醫師身分證字號，加密處理(現行編碼同)

急慢性病床天數

主、次診斷代碼(現行編碼同)，複數欄位

主、次手術處置(現行編碼同)，複數欄位

診察費(現行編碼為DIAG_DOT，診察點數)

病房費(現行編碼為ROOM_DOT，病房點數)

醫療費用(現行編碼為MED_DOT，醫療點數

部分負擔金額(現行編碼為PART_DOT

保險對象性別，現行編碼為ID_S

FEE_YM	APPL_TYPE	HOSP_ID	APPL_DATE	ID	ID_BIRTHDAY	GAVE_KIND	FUNC_TYPE	IN_DATE	OUT_DATE	E_BED_DAY	S_BED_DAY	PRSN_ID	DRG_CODE	TRAN_CODE	ICD9CM_CODE	ICD_OP_CODE	DIAG_AMT	ROOM_AMT	MED_AMT	PART_AMT	APPL_AMT	EB_APPL30_AMT	ID_SEX
200110	1	bcaaeaf9e4df941832f23098ad14acde05	20011113	012e61103b1db1b6e92ecf423de53aa5	19820921	4	2	20010922	20011002	10	0	fafce80d808c20064753940a911b6f97		3	483		2946	9080	29123	2912	26211	29123	M
200110	1	071df651df5c8286e464267a633ee47715	20011113	03f7ee2df48368ad0b33ddbe9d031bc6	19680415	3	3	20011023	20011026	3	0	e9a8af80c330b36569b976d41e8fc958		3	8501		828	2181	7843	784	7059	7843	M
200104	1	3084f92b53c894d5f4237a2c2b7fd6ce15	20010507	064a702137670c5c2f8c2d809a52d1af	19761226	4	3	20010330	20010406	7	0	872621bd44c5eb2b216794a60144fb50		3	6823		1656	4669	12941	1294	11647	12941	M
200107	2	6f03ebe2593d91fb32e6c488504f1d1912	20010810	1221516a5bd6cf717b05597337ff2742	19130527	9	AJ	20010509	20010724	21	0	2f279ae6ac0118dd46d3b267f7b15f9f		3	5579	4595	5892	43050	148255	0	127818	0	M
200105	1	6f03ebe2593d91fb32e6c488504f1d1912	20010611	1221516a5bd6cf717b05597337ff2742	19130527	3	6	20010426	20010507	11	0	20290058a9cc493dffaf0ca31906bfd22		3	82021	7935	3488	9108	59760	5976	53784	59760	M
200107	2	6f03ebe2593d91fb32e6c488504f1d1912	20010810	1221516a5bd6cf717b05597337ff2742	19130527	9	AJ	20010509		21	0	2f279ae6ac0118dd46d3b267f7b15f9f		2	5579	4595	5208	43050	165995	0	180600	0	M
200106	2	6f03ebe2593d91fb32e6c488504f1d1912	20010710	1221516a5bd6cf717b05597337ff2742	19130527	9	AJ	20010509		4	0	5c49060a65364abb24d7a68263aa0b55		2	486	9390	2124	25600	55405	0	55405	0	M
200106	1	6f03ebe2593d91fb32e6c488504f1d1912	20010711	1221516a5bd6cf717b05597337ff2742	19130527	4	AB	20010509		3	0	5c49060a65364abb24d7a68263aa0b55		2	486	9390	1027	8216	19425	0	19425	19425	M
200104	1	6f03ebe2593d91fb32e6c488504f1d1912	20010511	1221516a5bd6cf717b05597337ff2742	19130527	4	AB	20010402	20010406	4	0	01b3cd02229766b4e12651794f2fcd9e		3	4140	9399	1308	3312	9261	926	8335	9261	M
200109	2	b2e05507d85575a917d79ab8a6d91a5215	20011011	1221516a5bd6cf717b05597337ff2742	19130527	9	2	20010724		60	0	1e419ef4c120d225161d582d645e1bb8		2	486		12420	43620	207425	0	220000	100381	M
200111	2	b2e05507d85575a917d79ab8a6d91a5215	20011213	1221516a5bd6cf717b05597337ff2742	19130527	9	2	20010724	20011116	55	0	1e419ef4c120d225161d582d645e1bb8		6	5849		12028	39985	267134	0	225298	123520	M
200105	1	6f03ebe2593d91fb32e6c488504f1d1912	20010611	1221516a5bd6cf717b05597337ff2742	19130527	4	BA	20010509		13	0	fcadd21e2303ffc378404e92fec3af1a		2	41071	4576	5607	78000	263626	23000	240626	263626	M
200106	2	6f03ebe2593d91fb32e6c488504f1d1912	20010710	1221516a5bd6cf717b05597337ff2742	19130527	9	AJ	20010509		14	0	fcadd21e2303ffc378404e92fec3af1a		2	486	9390	5934	85600	201602	0	201602	201602	M
200102	1	6f03ebe2593d91fb32e6c488504f1d1912	20010312	1221516a5bd6cf717b05597337ff2742	19130527	4	AB	20010209	20010227	18	0	b7bc11a7bbd9203ab90a4c77e5530402		3	39891	8872	6208	66624	159749	15975	143774	159749	M
200112	2	6f03ebe2593d91fb32e6c488504f1d1912	20020108	1221516a5bd6cf717b05597337ff2742	19130527	9	AJ	20011116	20011210	24	0	2f279ae6ac0118dd46d3b267f7b15f9f		6	486	3927	13493	153600	406688	0	406688	406688	M
200112	1	c7ee9b9fcf923908d43cce5feac31356f04	20020111	14ea6889052401ada5dbac5f8846a5d9	20000719	4	CA	20011204	20011206	2	0	b17dbb630522a59d4f2ea8e31ed9022e		3	2392	8339	1065	2070	28973	0	28973	28973	F

(12) 外院診斷性及分期性手術處置
(13) 腫瘤分期
(14) 首次療程開始日期（包括決定不治療之日期，年月日）
(15) 首次手術日期（年月日）
(16) 原發部位確切手術切除日期（年月日）
(17) 外院及申報醫院原發部位手術方式
(18) 原發部位手術邊緣
(19) 放射治療儀器（如放射手術、近距放射治療、放射線同位素治療等 10 項分類）與治療開始、結束日期（年月日）、為放射治療原因（如禁忌症、個案死亡等 8 項分類）
(20) 體外放射治療技術（如二維或簡單電腦斷層定位、三維順形放射治療等 10 項分類）
(21) 全身治療開始日期（年月日）
(22) 化學治療開始日期（年月日）
(23) 荷爾蒙治療開始日期（年月日）
(24) 免疫治療開始日期（年月日）
(25) 骨髓/幹細胞移植或內分泌治療開始日期（年月日）
(26) 申報醫院緩和醫療（9 項分類）
(27) 首次復發日期（年月日）及復發形式；
(28) 最後聯絡或死亡日期（年月日）
(29) 死亡原因

5. 是否資料加密並無法得知相關資料，在 re-identification 技術下並無法完全確定去識別化之資料是否能在資料分析中維持
以醫療機構為例，資料檔中為加密處理，但若暫不論資訊科技之解密技術，搭配相關資料仍可大致猜出醫療機構所屬，並就此分析判斷個別保險對象是否至特定醫療機構就醫或住院
詳言之，醫療機構代號係加密處理，但資料檔中有醫事機構所在地（縣市鄉鎮區）與權屬別（分為公立與私立，而公立醫療機構包括部立及直轄市例醫院、縣市立醫院、公立醫學院校附設醫院、軍方醫院、公立機關(構)附設醫院、公立中醫院，而私立醫療機構包括醫療財團法人醫院、宗教財團法人附設醫院、私立醫學院校附設醫院、公益法人所設醫院、私立西醫醫院、私立牙醫醫院、私立中醫醫院）
則若個別保險對象之單行資料顯示其於台北市北投區（醫事機構所在地）之榮民醫院（權屬別）住院，則應可推測出該個案住院醫院為台北榮民總醫院

6. 須注意者為，衛福部科資中心尚有較為敏感之資料檔案，如低收入戶及中低收入戶資料庫、家暴通報明細檔、兒童及少年保護明細檔、性侵害通報明細檔等
這些資料因均屬衛福部所有資料，故不需再經所搜集資料之機關(構)同意，即可依申請程序進行資料檔串接
而以性侵害通報明細檔（Welfare-15: W_DPS_SEXAS）為例，其包括「被害人個人歸戶及個人資料比對的鍵值（ID_C）」與「相對人個人歸戶及個人資料比對的鍵值（ID_D）」，故可以性侵被害人 ID 或加害人 ID 為鍵值串接健保資料庫檔案，得知該個案被害人或加害人之就醫紀錄
而其中性侵害通報明細檔資料包括：
(1) 被害人與加害人之出生年月
(2) 被害人職業（18 類）
(3) 被害人教育程度（11 類）與就學狀況（7 類）
(4) 被害人國籍
(5) 被害人障礙別（33 類）
(6) 嫌疑人數
(7) 嫌疑人與被害人關係（包括配偶、前配偶、直系親屬、旁系親屬、家人的朋友、男/女朋友、前男/女朋友、同事、同學等 22 類）
(8) 案發場所（32 類）
(9) 案發縣市鄉鎮區

7. 除衛福部資科中心資料庫之資料檔外，依經其他機關(構)同意且符合若符合衛署醫字第 1010265083 號函規定，即可以相同之個人歸戶及個人資料比對的鍵值（通常為身分證字號），進行更多樣之資料串接。

3. 不同檔可以串連，並依譯碼簿所載之欄位進行串接，而此時前一點所說每一行所代表之每一位保險人之資料內容便將增加。
以衛福部資科中心之「全民健保處方及治療明細檔–西醫住院檔（HEALTH-02：H_NHI_IPDTE）」為例，譯碼簿說明「個人歸戶及個人資料對照鍵值：ID」，即表示以 ID（保險對象身分證字號欄位）進行串接，並可將不同資料檔與該保險對象有關之資料進行歸檔並整理。
以下便以不同檔之串接為例進行說明：
(1)「全民健保處方及治療明細檔_門急診–西醫、中醫及牙醫（HEALTH-01: H_NHI_OPDTE）」之「個人歸戶及個人資料對照鍵值」亦為 ID，因此便能與「全民健保處方及治療明細檔–西醫住院檔（HEALTH-02：H_NHI_IPDTE）」，而串接所得之資料將整合兩個檔案：
(A)除個別保險對象（ID 加密）之每次住院資料外（如前點所述，包括性別、醫師身分證字號(加密)及性別、就醫科別、入院日期、入出院年齡、住院天數(急性或慢性病床)、診斷代碼(ICD9-CM)、手術處置(ICD9-PCS)、就醫醫療機構代碼(加密)與權屬別及所在縣市鄉鎮區、給付類別、是否免部分負擔(包括免部分負擔之理由為重大傷病、分娩、低收入戶、榮民或遺眷家戶代表、多氯聯苯中毒油症患者)、依附就醫新生兒日齡、是否為汽車交通事故等）；
(B)亦能進一步得到同一保險對象之每次門診資料，內容包括但不限於以下項目：
a) 該次就醫之案件分類（包括西醫一般門診、西醫急診、西醫門診手術、西醫慢性病、洗腎、結核病、居家照護、精神疾病社區復健、安寧居家照護、職災案件、愛滋病防治替代治療計劃、愛滋病案件等，或牙醫一般門診、急診、門診手術、牙周統合照護、特殊專案醫療服務項目等，或中醫一般門診、慢性病、針灸、傷科、脫臼整復等）
b) 就醫科別（48 項分類）
c) 醫師身分證字號(加密)及性別
d) 就醫年齡
e) 特定治療項目
i) 包括癌症、內分泌代謝疾病、精神疾病、神經系統疾病、循環系統疾病、呼吸系統疾病、消化系統疾病、骨骼肌肉系統及結締組織疾病、眼及其附屬器官疾病、傳染病、先天畸形、皮膚及皮下組織疾病、血液及造血器官疾病、耳及乳突疾病等
ii) 進一步可分為特殊檢查（超音波、耳鼻喉科、內視鏡、病理組織、核子醫學、X 光、特殊造影、神經科檢查等 8 項分類）與特殊治療或處置（包括癌症放射線治療、癌症化學治療、高壓氧治療、眼科鐳射治療、血液透析治療、腹膜透析、根管治療、銀粉填充、牙周病手術、兒童斷髓處理、治療性牙結石清除、針灸、傷科治療、脫臼整復、腸病毒等共 129 項分類）
f) 國際疾病分類(ICD9-CM)
g) 主次手術
h) 給藥日份
i) 慢性連續處方箋總處方日數
j) 處方調劑方式
k) 就醫醫療機構代碼(加密)與權屬別及所在縣市鄉鎮區
l) 給付類別（17 項分類）
m) 是否免部分負擔(包括免部分負擔之理由為重大傷病、分娩、低收入戶、榮民或遺眷家戶代表、多氯聯苯中毒油症患者等，20 項分類)
n) 是否轉院、轉入醫療機構代碼(加密)與權屬別及所在縣市鄉鎮區。
(2)「全民健保重大傷病檔（HEALTH-08: H_NHI_CATAS）」之「個人歸戶及個人資料對照鍵值」亦為 ID，故串接後便可得到個別保險對象之重大傷病資料（包括罕病註記、健保署所定之重大傷病項目等）及單次就醫、住院資料

4. 除健保署內部資料檔之串接外，亦可與衛福部內部不同機關間之資料檔進行串接
如「癌症登記檔（HEALTH-14: H_BHP_CRFLF）」之資料來源來自國健署（非健保署），但仍屬衛福部資科中心之資料庫內容，而其「個人歸戶及個人資料對照鍵值」亦為 ID，與健保署之資料檔相同，故可串接；而串接後可依此得到個別保險對象（癌症患者）之及單次就醫、住院資料，以及該病患之相關癌症資料，內容包括但不限於以下項目：
(1) 身分證字號(加密)及性別
(2) 戶籍地（至縣市鄉鎮區，如高雄市苓雅區或苗栗縣通霄鎮）
(3) 癌症診斷年齡
(4) 個案分類
(5) 原發部位
(6) 側性
(7) 組織類型
(8) 分級/分化
(9) 癌症確診方式
(10) 腫瘤大小
(11) 區域淋巴結檢查數目及侵犯數目