個人資料去識別化過程驗證要求及控制措施

壹、目的

因應「政府資料開放」及「大數據」之推動，為驗證個人資料去識別化的過程，特訂定本驗證要求及控制措施。

貳、用語及定義

(1)個人可識別資訊(personally identifiable information, PII)：所有資訊其能用以識別此類資訊所涉之PII當事人，或係或得以直接或間接連結至PII當事人。

(2)識別(identification)：使用PII當事人所宣稱之屬性或所對其觀察之屬性，由一組個體中挑選出某特定個人之過程。

(3)去識別化(de-identification)：採取一組合理之步驟，移除識別資料與資料主體間之關聯的過程。

(4)直接識別資料(direct identifying data)：直接識別PII當事人之資料。直接識別資料係不需額外資訊或經由交互連結公開資訊中之其他資訊即可用以識別PII當事人之資料。例：身分證號碼、指紋等。

(5)間接識別資料(indirect identifying data)：僅於與其他間接識別資料一起使用時方足以識別PII當事人之資料。例：郵遞區號、生日、年齡等。

(6)匿名性(anonymity)：不允許個人可識別資訊(PII)當事人被直接或間接識別之資訊特性。

(7)匿名化(anonymization)：個人可識別資訊(PII) 不可逆地變更之過程， 以此方式，使得無法直接或間接識別PII當事人。

(8)匿名資料(anonymized data)：個人可識別資訊(PII)經匿名化過程輸出所產生之資料。

(9)擬匿名化(pseudonymization)：應用於個人可識別資訊(PII)， 以別名替換個人識別資訊之過程。(參照CNS 29100中「4.4.4擬匿名化資料」。)

(10)不可逆性(irreversibility)：由可識別至擬匿名之任何轉換的狀況，其由擬匿名追踪回原始識別符於計算上是不可行的。

(11)去連結資料(unlinkable data) ：僅包含難以由熟練的分析師以合理工作量連結至PII資料。

(12)重新識別(re-identification)：將已去識別化資料與原PII當事人重新建立關聯的過程。

(13)K-匿名性(K-Anonymity)：若發布之資料中所包含之PII當事人的資訊與至少K-1個人的資訊，無法區別。

(14)PII當事人(PII principal)：個人可識別資訊(PII)所關聯之自然人。

(15)PII控制者(PII controller)：判定個人可識別資訊(PII)處理之目的及方法的隱私權相關者，而非就個人目的使用資料的自然人。

(16)PII處理者(PII processor)：代表PII控制者並依其指示，處理個人可識別資訊(PII)之隱私權相關者。

(17)PII處理生命週期(PII processing life cycle)：包含PII之蒐集、移轉、使用、儲存、移除等階段。

(18)推論控制(inference control)：控制僅揭露無法據以推論出原PII當事人之資料。

(19) CNS 29191用語對應於本要求及控制措施之用語如下：

CNS 29191 用語	本要求及控制措施對應用語
宣稱者	PII 當事人
核發者	PII 控制者
指定開啟者	PII 控制者
查證者	(擬)匿名資料接受者

參、隱私權政策

要求事項(3.1.1)：涉及 PII 處理之組織的高階管理階層，應依營運要求及相關法律與法規，建立隱私權政策，提供隱私權保護之管理指導方針及支持。

控制措施(3.1.1.1)：隱私權政策應如下。

- 合於組織目的。
- 提供設定目標之框架。
- 包括滿足適用之隱私保全要求事項的承諾。
- 包括持續改善之承諾。
- 於組織內傳達。
- 公眾(或相關各方)可適時且容易取得。

控制措施(3.1.1.2)：組織應以書面載明其隱私權政策。

控制措施(3.1.1.3)：隱私權政策應依規劃之期間或發生重大變更時審查，以確保其持續的合宜性、適切性及有效性。

控制措施(3.1.1.4)：隱私權政策應依不同隱私權利害相關者，補充更詳細之 PII 處理規則及義務(例：特定部門或員工之程序)。此外，應載明於特殊設置下(例：存取控

制、告知條款、稽核等)，用以增強隱私權政策之控制措施。

控制措施(3.1.1.5)：內部隱私權政策應載明組織採用之目標、規則、義務、懲處規定、限制及/或控制措施，以滿足與其 PII 處理生命週期各階段有關之隱私保全要求事項。

控制措施(3.1.1.6)：組織應傳達予隱私權相關者下列資訊。

－ PII 控制者及所有相關之 PII 處理者的身分。

－關於移轉 PII 至 PII 處理者之政策。

－蒐集 PII 之目的。

－識別將蒐集之 PII。

－加強隱私權保護之作為及其目的。

－PII 當事人對其被蒐集之 PII 的法律權利。

控制措施(3.1.1.7)：外部隱私權政策應提供外部人員對組織隱私權實務作法聲明，以及其他相關資訊，如 PII 控制者之身分及辦公室地址、PII 當事人可能取得額外資訊之連絡窗口等。

控制措施(3.1.1.8)：組織應具備正式懲處過程，並傳達予員工及約用人員，以對違反隱私權者採取行動。

控制措施(3.1.1.9)：(告知之透明性)**若** PII 處理者非 PII 控制者，則 PII 處理者之隱私權政策應依循 PII 控制者之隱私權訂定。

控制措施(3.1.1.10)：組織若進行 PII 去識別化過程，則隱私權政策應包含下列項目，並應對外公布適宜之內容。

－敘明組織之去識別化作法，並以一般用語描述將使用何種去識別化技術。

－敘明備妥何種保護措施，以盡量減少可能之相關風險。尤其是，應敘明去識別化資訊是否會對外公開或僅有限揭露，及其公開原則(例：離群值之處理、K-匿名性之使用時機)。

－敘明對公眾開放關於進行之去識別化過程的任何風險。

－公開敘明關於公布已去識別化資訊之推理過程，說明如何衡量及取捨、考量或未考量哪些因素、原因為何。

實作指引：對公眾之資訊透通性可增加信任度，然基於資訊安全之考量，為不助長重新識別風險，組織應衡量是否需移除所公布之風險評鑑報告等文件中之某些資訊，或僅公布其彙總報告。

肆、PII隱私風險管理過程

要求事項(4.1.1)：組織應定期執行廣泛之 PII 風險管理活動並發展與其隱私保護有關的風險剖繪。

控制措施(4.1.1.1)：組織應建立 PII 處理生命週期各階段之風險管理過程。各階段應包含下列子過程。

- 建立全景過程：藉瞭解組織(例：PII 處理、職責)、技術環境及影響隱私風險管理之因素(亦即法規因素、契約因素、營運因素與其他因素)達成。
- 風險評鑑過程：藉識別、分析及評估 PII 隱私權原則之風險(可能有負面影響之風險)達成。
- 風險處理過程：藉定義隱私保全要求事項、識別及實作隱私控制措施以避免或減少 PII 隱私權原則之風險達成。
- 溝通及諮詢過程：藉從利益相關者得到資訊、對每一風險管理過程獲得共識，以及通知 PII 當事人與溝通風險及控制措施達成。
- 監視及審查過程：藉追查風險及控制措施，以及改善過程達成。

實作指引：隱私權衝擊評鑑為一項產出，其為風險管理之一部分，隱私權衝擊評鑑乃專注於確保遵循隱私權及資料保護法規之要求，以及評鑑於新的或大幅修改計畫或活動中的隱私權含意。

實作指引：隱私權衝擊評鑑宜框限於更大範圍之組織風險管理框架內。

伍、個人可識別資訊(PII)之隱私權原則

本節係管理個人資料之一般要求及控制措施，組織可參照 CNS29100 第 5 節之各項隱私權原則，實作其個人資料管理之作為。

陸、PII去識別化過程

<u>要求事項(6.1)</u>：組織應建立有效且周延之 PII 去識別化過程的治理結構。

控制措施(6.1.1)：組織應指定足夠數量具技術與法律知識之員工或約用人員進行 PII 去識別化。並應指定資深員工，負責授權及監督 PII 去識別化過程。此負責人員應有能力負責 PII 去識別化主要決策、宣達及協調組織之 PII 去識別化作法、召集組織內部及外部相關專家，並應能協助高階管理階層決定已去識別化資料之適當揭露形式(亦即公開或有限存取)。

控制措施(6.1.2)：應經由人員訓練，使 PII 去識別化工作人員清楚認識 PII 去識別化技術、所涉及之所有風險及減輕此等風險之措施。尤其是，各工作人員應了解其於確保安全進行去識別化之特定角色。

控制措施(6.1.3)：應提供獨立及隔離(無法連線)空間(及系統)進行 PII 去識別化工作，並管制及記錄人員與資料之進出(及存取)，且人員不得攜帶任何具照像及記錄功能之設備進入工作區域。

控制措施(6.1.4)：應備妥以文件記錄之程序，識別決定是否對資料進行 PII 去識別化、其實施方法，以及產生之資料是否需揭露、揭露原則及揭露方式。

控制措施(6.1.5)：組織應備妥以文件記錄之程序，用以識別於實務上去識別化可能是有問題或難以達成之情況。例：難以評估重新識別之風險，或是對某些個人之風險太高。

控制措施(6.1.6)：組織應依據法律規定、組織任務、營運要求、資料使用對象及目的、所持有包含 PII 之資料內容、型式及數量、資料揭露範圍、處理成本及風險評鑑結果等因素，選擇適宜之去識別化方法，並經管理階層核准，且以文件記錄。

控制措施(6.1.7)：資料去識別化過程應留下紀錄、全程受監督(例：全程錄影)，且其紀錄應無法竄改，並定期稽核。

控制措施(6.1.8)：委外處理 PII 去識別化時，組織應監督及監視委外處理活動。原始資料以不攜出組織場域為原則。含有 PII 之資料應經組織之高階管理階層核

准方可攜出場域外，而受委託單位須依組織之隱私權政策及隱私權原則妥善並安全保存原始資料，並於完成 PII 去識別化後，立即歸還組織或安全銷毀。

要求事項(6.2)：組織高階管理階層應監督及審查 PII 去識別化過程之治理的安排。

控制措施(6.2.1)：組織應管理，關於 PII 去識別化之任何新指引、法規、法律、裁判、行政解釋、可用技術或威脅之相關知識，並據以評估風險。

控制措施(6.2.2)：組織應與同行業或從事類似工作之其他組織分享並交流關於 PII 去識別化之知識。

控制措施(6.2.3)：組織應定期進行隱私衝擊評鑑(privacy impact assessment, PIA)，並應公布其 PIA 報告，顯示如何處理風險評鑑過程。PIA 應包含所採用去識別化技術之有效性，以及評估重新識別風險，以制定風險緩解措施。

實作指引：專家用以協助決定資料可識別性風險之原則如下。

原則	解釋	風險值範例
重現性 (Replicability)	根據資料將持續連結至PII主體之機率，將資料屬性定出風險等級之優先序。	**低**: 病患之口腔疾病，會改變。 **高**: 病患之牙齒照片相對穩定。
資源可用性 (Resource Availability)	判定哪些外部資源含有特定個人之識別資料及資訊中之重覆特性，以及何人被允許存取該資源。	**低**:實驗室報告中之個人身份通常不會對實驗室外披露。 **高**:個人身份及人口資料往往出現於公共資源中，例:出生、死亡及婚姻狀態。
區別性 (Distinguish)	判定某PII資料可於資料集之中區別的程度。	**低**:據估計在美國使用出生年、性別及郵遞區號前3碼之組合約有0.04%機率可唯一識別某居民。此意指僅經由此等資料之組合可識別特定居民。 **高**:據估計在美國使用出生日期、性別及5碼郵遞區號可唯一識別某居民之機率超過50%。此意指經由此3個資料之組合可識別一半以上的美國人。
評鑑風險(Assess Risk)	重現性風險、資源可用性風險及區別性風險越高，被識別之風險越高。	**低**:資料不具區別性，但其可能並未獨立重現，且很少於許多人可存取之多個資源中揭露。 **高**:人口資料具高度區別性、高度重現性並揭露於公共資源中。

可使用此等原則，判定資料集之中的PII資料之風險值。

控制措施(6.2.4):組織之高階管理階層應決定已移除 PII 之資料之可接受剩餘風險。

控制措施(6.2.5)：管理階層應依規劃之期間或發生重大變更時審查去識別化過程。

控制措施(6.2.6)：告知之透通性，同控制措施(3.1.1.9)。

控制措施(6.2.7)：組織應依據對所收到回饋之分析，持續且及時審查 PII 去識別化過程。審查時應使用〝重新識別測試〞技術評鑑重新識別風險及減緩風險之措施。

控制措施(6.2.8)：應對已移除 PII 之所有資料進行系統化(自動或人工)檢查，確保其中未包含直接識別資訊，以及非必要保留之間接識別資訊。並確保必要保留之間接識別資訊皆已(經由匿名化、擬匿名化或其他方法)合理去除與 PII 當事人之連結。政府「開放資料」，須達「匿名資料」或「不可逆之擬匿名資料」之程度。

要求事項(6. 3)：組織應訂定如下之 PII 去識別化步驟，並依此進行去識別化。

步驟 1：由領域專家(或有經驗人員)判定資料集之中的直接識別資料。

步驟 2:將直接識別資料遮罩(或變換)。即移除直接識別資料或將其(擬)匿名化。

步驟 3：建立威脅模型。組織分析可能使用額外資訊或間接識別資料進行重新識別攻擊之各種情境，判定各種“可能威脅”。

步驟 4：判定最小可接受使用之資料。於此步驟中，組織確定 PII 去識別化資料之用途，並據以判定可能需去識別化之資料的最大數量。

步驟 5：使用步驟 3 所建立之模型，確定重新識別攻擊之風險的檻值。組織判定使用所有已去識別化資料之可接受風險，以及可能降低風險之各項控制措施。

步驟 6:由來源資料庫中取得(樣本)資料。因資料庫可能很巨大，故取樣測試，以協助規劃。

步驟 7：評估實際之重新識別攻擊風險。即計算實際被重新識別之風險。

步驟 8:比較實際之被重新識別的風險與檻值。即比較步驟 7 與步驟 5 之結果。

步驟 9：設定參數並套用至所有需去識別資料。若實際風險小於最小可接受風險，則套用去識別參數，並變換資料。若風險過高，則需考慮新的參數或變換。

步驟 10：對解決方案進行診斷。測試已去識別資料，以確保其具有足夠效用，並確認於允許之參數範圍內，重新識別攻擊是合理的不可能的。

步驟 11：輸出已去識別資料至外部資料集。最後，輸出已去識別資料，並將所使用之去識別技術、參數、威脅模型、風險值及各項相關資料，記錄於書面報告中。

控制措施(6.3.1)：應對待移除 PII 之資料集，先行備份，必要時應依隱私權原則進行前置處理，抽出最少需處理之資料、欄位或其部分。

控制措施(6.3.2)：應依待移除 PII 之資料型式(例：書面文字資料、書面圖片、文字檔、資料庫、圖片檔等)，選擇適當去識別作法及工具。

控制措施(6.3.3)：組織應依資料揭露(公布)對象及資料敏感性，設定推論控制之檻值(例：K-匿名性之最小 K 值、揭露筆數占整體筆數之最小百分比)。不得揭露超過檻值之資料。

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

實作指引：推論控制通常使用之「N 回應 k%支配」規則，指若超過 k%以上之揭露(或公布)資料係來自少於 N 筆資料，則不得揭露(或公布)該等資料。例：不應公布比爾蓋茲所住社區之家庭平均年收入，因其年收入占整個社區之年收入比例太高，若公布此資料，易被用以推估出比爾蓋茲之年收入。

實作指引：PII 資料去識別化技術範例如下。

－修訂或移除 PII 資料：使 PII 資料成為(人類或電腦)不可視。例：用黑筆將書面資料塗黑、將人臉打馬賽克、移除檔案中 PII 資料、將資料加密、移除整筆資料等。

－模糊化 PII 資料：於資料中加入隨機〝雜訊〞。例：將某筆個人資料之年齡加 5 歲，而下 1 筆資料之年齡加 8 歲、只提供對部分資料之統計數字。

－概化 PII 資料：降低資料之精確度，使其較不特定。例：將年齡 25 歲，變

成年齡 20~29 歲。

－合併 PII 資料：將數個 PII 資料項合併成一資料項，使其較不具敏感性。例：2014 年及格人數為 2 人，2015 年及格人數為 3 人，合併為 2014~2015 年及格人數為 5 人。

－以平均值置換 PII 資料：例：將某筆個人資料之年齡以所有資料之平均年齡置換。

－一致性置換 PII 資料：將所有 PII 資料位移相同量，以保持資料間之關係。例：將所有個人資料之年薪均加 10 萬元。

－交換 PII 資料：將 2 筆資料之 PII 資料欄位內容交換。例：將 2 筆資料之年齡交換。

實作指引：K-匿名性為去識別化後判定是否揭露資料之限制條件，組織應僅揭露具相同值筆數大於等於 K 之資料。增加 K 值可降低資料遭重新識別之風險。K-匿名性可與各種去識別化技術一起使用。

要求事項(6. 4)：組織應對 PII 遭非預期揭露備妥災難復原計畫。

控制措施(6.4.1)：應及時回應來自自認為個人資料遭揭露民眾之申述及查詢，並依已建立之程序採取因應措施。

控制措施(6.4.2)：應備妥程序，因應公開資料遭重新識別而揭露個人隱私之情況，包含：移除可能揭露個人隱私之資料，重新處理；停止或修改(採取更嚴格之)去識別化過程。

控制措施(6.4.3)：當公開資料遭重新識別而揭露個人隱私時，應告知隱私遭揭露之個人，並協助其採取必要之彌補措施。

要求事項(6. 5)：組織應備妥程序，對已移除 PII 之資料，依可接受風險，定期進行〝重新識別測試〞。

控制措施(6.5.1)：應對已移除 PII 之資料進行〝重新識別測試〞，包含：

－ 搜尋網頁，嘗試連結 PII 當事人。

－搜尋全國或地方新聞資料庫，嘗試連結 PII 當事人。

－搜尋政府單位或其他組織之開放資料，嘗試連結 PII 當事人。

－以社群網路嘗試連結 PII 當事人。

控制措施(6.5.2)：因公眾可用之資料庫，隨時會增長，故應定期重新對已移除 PII 之資料進行〝重新識別測試〞，以重新評鑑其風險。

柒、重新識別 PII 之要求 (此部分為選項)

說明:重新識別係將已去識別化資料與 PII 當事人重新建立連結之過程。此將增加去識別化過程之複雜度。組織需重新識別 PII 當事人之理由可能包括：

－對資料完整性之檢驗。

－檢查是否有疑似重複之資料。

－加入新資料。

－連結至額外研究變量。

－符合性稽核。

－重大發現需通知 PII 主體或相關單位。

－進行後續進一步研究。

－法律要求。

組織有重新識別需求時，應符合 CNS 29191 所有要求事項。

要求事項(7.1)：經匿名(或擬匿名)處理後資料之接收者應僅能鑑別 PII 當事人之資料屬性，而無法識別出 PII 當事人。

控制措施(7.1.1)：經匿名(或擬匿名)處理後之資料不得提供任何可用以識別出資料之 PII 當事人，但必要時可允許資料接收者查證經匿名(或擬匿名)處理後之資料(或其屬性)是否真實。

要求事項(7.2)：同一 PII 當事人之經匿名(或擬匿名)處理後之不同資料，不得提供具有聚合後能連結至該 PII 當事人之資訊。

控制措施(7.2.1)：資料接收者取得之經匿名(或擬匿名)處理資料，不得包含可據以連結 PII 當事人之間接識別資料。

要求事項(7.3)：資料經可逆之擬匿名處理後，應可由 PII 控制者重新識別 PII 當事人。

控制措施(7.3.1)：PII 控制者應備有重新識別 PII 之程序，規定所使用方法、所需資訊、授權及啟動流程。

控制措施(7.3.2)：應定期審查重新識別 PII 之程序的有效性。

控制措施(7.3.3)：為使 PII 控制者之後能重新識別 PII 當事人，將資料經可逆之擬匿名處理後產生之紀錄單，應提供足以識別 PII 當事人之必要資訊。

控制措施(7.3.4)：PII 控制者對將 PII 資料經可逆之擬匿名處理所產生之紀錄單及重新識別所需之必要資料，應妥善加密，持續保存。

實作指引：於適當情況下， PII 控制者可使用其他資訊以重新識別 PII 當事人。

要求事項(7.4)：PII 控制者應提供能正確重新識別 PII 當事人之證據。

控制措施(7.4.1)：為避免 PII 控制者之不誠實宣稱， PII 控制者應提供正確履行重新識別 PII 當事人之程序的證據。

控制措施(7.4.2)：重新識別 PII 當事人資料之過程應留下紀錄、全程受監督(例：全程錄影)，且其紀錄應無法竄改。

參考資料

[1] 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

[2] CNS 29100 資訊技術－安全技術-隱私權框架

[3] CNS 27001 資訊技術－安全技術－資訊安全管理系統－要求事項

[4] ISO/IEC 27018:2014 Information technology－Security techniques－Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors.

[5] ISO/IEC 29101:2013 Information technology－Security techniques－Privacy architecture framework.

[6] ISO/TS 25237 :2008－Health informatics－Pseudonymization.

[7] Simson L. Garfinkel, De-Identification of Personally－Identifiable Information, DRAFT NISTIR 8053 1, 2015.

[8]Anonymisation: managing data protection risk code of practice, ico, UK, 2012. *https://ico.org.uk/media/1061/anonymisation－code.pdf*

[9]IHE IT Infrastructure Technical Committee, IHE IT Infrastructure Handbook: De-Identification, 2014.

[10] Bradley Malin, A De-identification Strategy Used for Sharing One Data Provider's Oncology Trials Data through the Project Data Sphere® Repository, 2013.