論述 》管理 · 資訊

行政院主計總處個資去識別化作業辦理情形與成效

為落實個人資料保護及管理，行政院主計總處以該處地方統計推展中心之家庭收支調查資料，及國勢普查處 104 年農林漁牧業普查資料為驗證範圍，通過財團法人台灣電子檢驗中心之個資去識別化過程驗證，特撰文簡要說明行政院主計總處之辦理情形與成效。

王興娟（行政院主計總處主計資訊處分析師）

壹、前言

行政院主計總處（以下稱本總處）爲落實個人資料（以下簡稱個資）之保護與管理，促進個資之合理利用，於 102 年度已建置「個人資料管理制度（PIMS）」。行政院爲促進公務機關釋出更多符合個人資料保護法要求之原始資料供民間加值應用，請經濟部研訂相關認證標準，經濟部標準檢驗局於 103 年 6 月發布 CNS29100（訂定隱私權框架之標準）及 104 年 6 月發布 CNS29191（訂定部分匿名及部分去連結之標準），作爲我國現階段推動個人資料去識別化之驗證標準。

本總處爲落實個人資料保護及管理，105 年以本總處地方統計推展中心之家庭收支調查資料爲驗證範圍，推動個資去識別化過程驗證作業，106 年擴大驗證範圍，加入國勢普查處 104 年農林漁牧業普查資料，通過財團法人台灣電子檢驗中心（Electronics Testing Center，Taiwan； 簡稱 ETC）驗證。

貳、本總處個資去識別化過程作業辦理情形

一、驗證程序概述

一項管理作業程序若要通過第三方驗證，在確認驗證範圍後，需建置一系列符合 PDCA 之循環式管理活動，即由 P（Plan） 計畫、D（Do）執行、C（Check） 查核及 A

（Act）行動等四大步驟過程所構成一連串追求改善精進之目標管理。以第三方驗證程序而言，大致包括下列幾項：

（一）盤點清查驗證範圍資料。

（二）進行風險評鑑作業，擬訂可接受風險值。

（三）擬訂風險處理計畫。

（四）建立四階程序書。

（五）進行內部稽核作業。若有稽核發現，需擬訂內稽矯正預防處理作業。

（六）進行管理審查委員會。

（七）進行第三方外部驗證作業，若有稽核發現，需擬訂外稽矯正預防處理作業。

（八）次年回到步驟（一），檢視資料是否有更新，及作業是否有需要改善精進之處。

二、本總處個資去識別化過程作業

本總處個資去識別化過程作業範圍，選定本總處地方統計推展中心之家庭收支調查資料，及國勢普查處 104 年農林漁牧業普查資料，依前述驗證程序說明如下：

（一）盤點清查驗證範圍資料

本階段應了解家庭收支調查資料及農林漁牧業普查資料之調查方法、抽樣方式，最後取得哪些調查欄位資料及資料筆數。有關個資去識別化最重要的一項工作爲依蒐集資料特性判定資料集之中的直接及間接識別欄位。間接識別欄位意指僅以該資料不能直接識別，須與其他資料對照、組合、連結等，始能識別該特定之個人。若無法判定間接識別欄位，會有資料洩漏的風險存在。

（二）執行個資去識別化程序

1. 家庭收支調查資料

(1) 由承辦人員依資料特性判定資料集之中的直接及間接識別欄位。

(2) 將原始資料存放於不對外連線之電腦保存；將待執行去識別化作業之資料使用隨身碟由專人送至個資去識別化專屬工作區域。

(3) 在專屬工作區域中，使用個資去識別化工具進行個資去識別化作業。

(4) 個資去識別化方式：

i. 將戶號轉爲流水號（共 8 碼）。

ii. 將敏感且無開放目的之欄位抹除爲 0。

iii. 資料經去識別化後，由專人檢查，確認資料皆已依照程序執行去識別化作業。

2. 農林漁牧業普查資料

(1) 由承辦人員依資料特性判定資料集之中的直接及間接識別欄位。

(2) 將原始資料存放於專案專用硬碟中保存；將待執行去識別化作業之資料由專人送至資料管制室。

(3) 在資料管制室使用個資去識別化工具進行個資去識別化作業。

(4) 個資去識別化方式：

i. 將村里代號、普查區、連續編號等欄位採遮罩處理。

ii. 將敏感且無開放目的之欄位抹除爲0。

iii. 資料經去識別化後，由專人檢查，確認資料皆已依照程序執行去識別化作業。

3. 執行本項作業，應特別注意以下重點

(1) 前述二項資料集依資料集特性，選取去識別方式，如：隱碼遮罩、區間概化取平均值或中間值等方式，進行資料概化。接續依選取之演算法（如：K、L、T 演算法），篩出最適當之策略進行去識別化。愈多間接欄位進行資料概化，資料去識別化後之資料品質將愈粗糙，同時影響之後資料研究成果。

(2) 本總處調查資料欄位眾多，間接欄位高達數十個，需配合去識別化工具協助處理，目前採用最常使用之 K 匿名法（K－Anonymity），隨著調整去識別方式，可立即進行資料處理並計算出 K 值，同時計算出資料集被再識別出之機率；藉由工具協助反復調整去識別方式並兼顧資料之品質。

（三）進行風險評鑑作業

依本總處個資去識別化風險評估及管理程序書並配合使用個資去識別化工具進行風險評鑑作業。

1. 執行風險評估程序

(1) 依據個人資料作業流程衝擊分析之結果，進行風險評估，並將其登錄於「個人資料去識別化作業風險評估表」，計算對應之風險值後，訂定可接受風險。

(2)「個人資料去識別化作業風險評估表」包含資料集名稱、衝擊值、重新識別機率及風險值等欄位。

i. 重新識別機率（RIR）：GR×IR×AIR×NPR，若 NPR 爲 0，則不列入計算。

(a) 群組率（GR）：係指資料集進行去識別化作業分群後，群組數除以資料總筆數的值，群組率越小，代表平均 K 值越高，資料遭重新識別機率越低。

(b) 最高識別率（IR）：係指找出資料量最少的群組，取其數量之倒數。

(c) 可接受最高識別率（AIR）：係指資料控制者須自行定義的可接受值，該值依資料欄位屬性及資料量大小不同而定，如資料控

制者僅要求K值>=2，則可接受最高識別率則可定爲0.5。

(d) 未通過可接受最高識別率資料比例（NPR）：係指所有低於最高識別率群組資料量總合除以資料總筆數。

ii. 風險值係運用「個人資料衝擊分析表」針對個人資料作業流程進行衝擊構面評估之衝擊值，再考量該資料集進行去識別化作業後，去識別化之資料集被重新識別的機率成爲風險值。

iii. 風險值之計算方式：個資作業流程衝擊值×重新識別機率。本總處個資去識別化資料集進行風險評估後之結果如表1。

2. 執行風險管理程序

(1) 可接受風險值之計算方式係運用個人資料衝擊值範圍之中位數，再考量去識別化資料之開放方式，透過去識別化工具設定K門檻值，計算K門檻值之最大重新識別機率成爲可接受風險值。

(2) 去識別化資料之開放方式若爲「約定使用」，則K門檻值設定爲3；若爲「公開使用」，則設定爲20。K值係指資料集進行去識別化作業分群後，所有群組中資料量最少者（或稱最小群組）之數量。

(3) 可接受風險值係爲個人資料衝擊值範圍之中位數×K門檻值之最大重新識別機率。

(4) 經本總處核定可接受風險值，針對高於可接受風險值之風險則須進行風險處理。

（四）重新識別機率驗證

1. 爲了解資料集於去識別化作業後資料分布的狀況，有別於風險評鑑，本總處另採用重新識別機率驗證的方式，透過動態調整「可接受最高識別率」，

表1　個資去識別化資料集之風險評估結果

個資去識別化資料集名稱	重新識別機率	風險值	可接受風險值	風險評估結果
訪問調查結果（家庭收支資料集）	2.1992%	0.3079	0.0033	高於可接受風險值
普查資料（農牧業）	2.0795%	0.2911	0.0001	高於可接受風險值
普查資料（農事服務業）	11.1111%	1.5556	0.0109	高於可接受風險值
普查資料（林業）	5.3927%	0.7550	0.0005	高於可接受風險值
普查資料（漁業）	4.7036%	0.6585	0.0003	高於可接受風險值

資料來源：作者自行整理。

論述 》管理 · 資訊

觀察「未通過可接受最高識別率資料比例」。本總處個資去識別化資料集進行重新識別機率驗證後之結果如表 2。

2. 針對「約定使用」且未通過重新識別機率驗證之資料集，應依本總處「個資去識別化作業程序書」之規定，僅限於學校、學術機構、政府機關或民意機關等申請，但有特殊情形者可經主管同意後釋出。

（五）建立四階程序書

本總處於 103 至 104 年整併資安及個資之四階程序書，爲推動個資去識別化作業，另訂定符合去識別化之程序規範，重要程序規範包含統計調查隱私權政策、個資去識別化作業程序書、個資去識別化風險評估及管理程序書、個資去識別化管理制度適用性聲明等，此階段申請提出驗證之程序規範計有一階程序書 2 份、二階程序書 14 份、三階程序書 1 份、四階程序書 8 份。

（六）進行內部稽核作業及管理審查委員會

這部分作業併同本總處資安暨個資推動業務一併辦理，若有稽核發現，需擬訂內稽矯正預防處理作業。

（七）進行第三方外部驗證作業

目前國內唯一政府認可執行個資去識別化過程驗證單位爲財團法人台灣電子檢驗中心（Electronics Testing Center，Taiwan； 簡稱 ETC），故備齊相關資料向 ETC 申請，於管理審查委員會完成後進行外部驗證作業，並通過驗證取得證書如下頁附圖。

表 2　個資去識別化資料集之重新識別機率驗證結果

個資去識別化資料集名稱	可接受最高識別率（AIR）	未通過可接受最高識別率（NPR）	未通過可接受最高識別率（NPR）之標準	重新識別機率驗證結果
訪問調查結果（家庭收支資料集）	25.1%	7.6045%	15%	通過
普查資料（農牧業）	33.4%	0.0000%		通過
普查資料（農事服務業）	25.1%	100.0000%		未通過
普查資料（林業）	25.1%	14.8469%		通過
普查資料（漁業）	8.4%	61.8059%		未通過

資料來源：作者自行整理。

參、本總處個資去識別化過程作業辦理成效

一、增進民眾對本總處之信任感

本總處依統計法舉辦之統計調查，受調查之當事人應依據統計法第 15 條規定，均應依限據實答復。相關調查、統計人員對各種統計調查取得之個別資料，應予保密，並如實登載。爲增進對本總處調查資料集之良善保護管理責任，調查資料於報告公告前，均會進行個資去識別化作業，以增進民

眾對本總處舉辦調查之信任感及配合度。

二、調查資料集預先進行風險管理

針對已去識別化之資料進行風險評鑑，相當於預先進行風險管理，同時可預估各式可能風險情境，並對較高機率發生之風險情境事先擬訂風險處理計畫，俾利事情發生後快速並完善的解決。

三、促進政府開放資料及大數據應用

針對已去識別化之資料，因已無法經由任何方法識別出特定個資資料，則該資料不適用個人資料保護法，是保障隱私，又能進行大數據分析之最適實務方案；且去識別化作業需嚴謹，澈底切斷資料內容與特定主體間之連結。

肆、結語

本總處是第2個通過個資去識別化過程驗證之公務機關，有關本總處之調查公告資料，為讓民眾信任放心，本總處會善盡個資保護責任，均能符合個資去識別化過程規範，以作為政府機關之表率，並期許帶動其它政府機關，或國內如金融業、科技業等持有巨量民眾個資的產業，對於個人資料進行保護，進而支持政府開放資料及大數據應用之推動。

❖

附圖　本總處個資去識別化過程驗證證書

個人資料去識別化過程驗證證書

ETC

茲 證 明

行政院主計總處

地址：臺北市忠孝東路1段1號

經本中心稽核結果符合 CNS29100
資訊技術-安全技術-隱私權框架

參考「個人資料去識別化過程驗證要求及控制措施」

予以認可登錄，範圍如下：

1. 行政院主計總處家庭收支調查資料
   地址：南投市中興新村光明路25號(中部辦公園區)
2. 行政院主計總處農林漁牧業普查資料
   地址：台北市中正區廣州街2號

另包含廠區範圍名稱：
廣博大樓1樓機房(地址：台北市中正區廣州街2號)

原始登錄日期　：中華民國 105 年 12 月 02 日
發 證 日 期　：中華民國 106 年 11 月 15 日
有 效 期 限　：中華民國 108 年 12 月 01 日
驗證證書編號　：1PII002-01

ETC 財團法人台灣電子檢驗中心

執行長 李海清

財團法人台灣電子檢驗中心地址：桃園市龜山區文明路29巷8號　　頁數：第1頁/共1頁

資料來源：財團法人台灣電子檢驗中心核發。