



# 行政院主計總處普(抽)查資料供應方式

因統計調查取得之資料涉及受訪者隱私，爰依統計法及資通安全法規，分級管理各項普抽查資料，並研訂資料供應方式，以維護資料安全及多元應用。

陳嘉鴻 (行政院主計總處國勢普查處專員)

## 壹、前言

行政院主計總處(以下簡稱總處)辦理各項普(抽)查作業，蒐集大量社經資訊，並於維護資安前提下，致力於資料的推廣應用。關於調查資料之釋出，凡經彙整統計的次級資料，因無個資疑慮，已於總處全球資訊網及三大普查統計資料平台提供各界查詢應用；至調查之個別資料，須經適當統計方法及資訊技術處理後，在無暴露個別資料疑慮下，分級提供各界應用。近年來民眾對個資保護觀念日益提升，總處爰依統計法、個資法及資通

安全管理法等規範下，研訂普(抽)查資料供應原則及作業規範，以維護資訊安全及強化資料應用價值。

## 貳、主要國家普查個別資料供應方式

近年來，面對開放資料(Open Data)的洪流，各主要國家的政府統計機構，積極強化調查資料供應方式，期提供多面向服務，回饋各方使用，進而提升各界對統計調查參與程度。惟依統計法及聯合國官方統計基本原則略以，統計調查取得之個別資料應予保密，是以，如何兼顧資料安全與資料應用

價值，為各國統計機關共同之課題。目前各國普查個別資料供應方式，依資料機敏程度及使用對象，大致可歸納為以下 5 種：

### 一、公開提供

普查資料經去識別處理後，以分層變數進行抽樣、關聯分析與檢定，並配適樣本權重，產製公眾使用個別資料樣本檔(Public Use Microdata Sample Files, PUMS)，使抽樣檔可推計至全檔總數及關鍵變數間之關聯，公開提供各界免費使用，增進分類組合彈性應用。處理情形包括兩部分：

#### (一) 變數處理

為降低個別資料被揭露風險，依變數性質先進行去識別處理，主要方法有：

1. 資料置換（Data Swapping）：將鄰近相似屬性個體進行變數值交換。
2. 頂級編碼（Top Coding）：對特定範圍或超過門檻值之資料，其數值歸至某特定類別。
3. 排除特定資料：對於地理區、分類或變數值設定門檻，未達者不予提供。如美國人口普查對地理區或種族，不提供人口數少於 8,000 人者；韓國則排除外國人及非普通住戶等。
4. 年齡擾動（Age Perturbation）：家戶人口數較多時，調整戶內成員年齡以保護隱密性。
5. 插補資料：將極端值改以插補值取代，但須注意變數間可能造成之影響。
6. 合併類別：類別次數未達最低門檻者，將相近類別合併計數。

## （二）產製抽樣檔

一般以地理區及屬性分類等特性進行抽樣設計，產製 1% 至 5% 的個別樣本檔，並

置於網站提供免費下載使用（附表）。例如美國從全國抽取 1%、州級別抽取 5% 不重複的人口普查長表樣本檔；英國抽取 1% 人口普查樣本檔；澳洲以人口普查區為單元，考量地理區、家戶及個人等，抽取 1% 樣本，產製基本保密資料檔（Basic Confidentialised Unit Record Files, CURFs）；韓國從全國抽取 1%、區域別抽取 2% 之兩類樣本檔。

## 二、申請後提供

此類資料經去識別處理及抽樣作業後，保留較多的變數內容，致使用上較為嚴謹，須申請後提供。例如英國抽取 5% 人口普查樣本檔，須經申請並簽具保密切結書；日本抽取 1% 人口普查樣本檔，僅供機構、團體或民營企業，須符合公共利益（教學、改善人民福祉、推廣統計等）經申請後計費使用。

## 三、遠端連線作業

使用者經申請後利用網路遠端連線統計機構伺服器，模擬終端機在該環境中處理資料，各項作業及輸出結果均受系統的監控管理，且所用資料均經過去識別處理，在保護資料隱私及安全下，增進資料取用的便捷性。

澳洲從全國抽取 5% 人口普查樣本，產製擴展保密資料檔（Expanded CURFs），更具詳細的資料，須經申請後利用遠端存取資料實驗室（Remote Access Data Laboratory, RADL）使用，該環境提供常用統計分析軟體應用，但抑制部分指令，避免瀏覽底層資料，具自動化檢核機制，禁止輸出具識別個資風險之結果；英國從全國抽取 10% 人口普查樣本檔置於虛擬微觀資料實驗室（Virtual Microdata Laboratory, VML），須申請後經由雲端帳

附表 主要國家公開提供普查個別資料抽出比率

國家	美國	英國	澳洲	韓國
抽出比率	全國 1% 州級別 5%	1%	1%	全國 1% 區域 2%

資料來源：各國統計（普查）局網站。

## 論述》統計·調查



戶登入進行資料處理，僅可輸出檢核過的分析結果。

### 四、管制室作業

由使用者提出申請，親赴機關指定之資料管制室作業，在實體隔離的環境下，使用資料均無法直接辨識個別資料，攜出結果須符合申請目的且經檢查未涉及個別資料，採行國家如美國、英國、澳洲、日本、韓國及我國等。

### 五、客製化統計編製

使用者申請需用之統計項及分類，由統計機構衡酌人力後，專門客製產生相關統計表件，並經審查無暴露個別資料疑慮後提供，一般會依其處理作業複雜度收取費用，採行國家如美國、澳洲、日本及韓國等。

### 參、我國普抽查資料供應情形

為提升普抽查資料應用價值並保護調查個別資料隱私，總處於 98 年訂定「提供普（抽）查資料管制作業要點」，依資料被揭露風險程度分為二類，第一類包含原始資料經同

類合併、彙整等處理之次級資料，以及去除個別識別碼及較具敏感性資料特徵值後之原始資料，可由資料使用者提出申請，經審核後提供電子檔。第二類指去除個別識別碼之資料，必須於總處資料管制室臨場作業，學校、學術研究或其他專業機構攜出資料僅限彙總之統計結果，且須經總處人員檢查後提供。

為保護調查個別資料安全並兼顧其應用價值，是各國統計機構資料供應精進的共同課題。面對技術方法的提升，以及對微觀資料需用的殷切與擴增，總處檢視現行供應作業方式與規範，並參酌各主要國家作法，相應調整普（抽）查資料供應方式，依資料被揭露風險程度擴增分為四類，以滿足更多不同目的及身分的使用者應用，各類說明如下：

#### 一、公開提供

新增之供應方式，以家戶面為基礎之農林漁牧業普查與人口及住宅普查產製 1% 及 5% 的抽樣檔，提供各界自行上網下載使用；至工業及服務業普

查因業別集中程度高，抽樣檔將大幅減損資料應用價值，爰不予提供。

104 年農林漁牧業普查抽樣檔已於去（108）年 10 月上線供應，係以合併類別及排除敏感性資料或個體等方法，進行關鍵變數去識別處理，並利用變數進行分層抽樣，配適各戶權重，可推計至全檔總數及關鍵變數間之關聯；依各業家數多寡提供抽樣檔，其中農牧戶及林戶抽取 1%，獨資漁戶抽取 5%。

#### 二、申請後釋出

申請後釋出提供經同類合併、彙整等處理後之普查次級資料，包括農林漁牧業與工業及服務業普查村里檔，以及去除個別識別碼及較具敏感性資料特徵值後之抽樣調查資料。

#### 三、管制室作業

管制室作業提供去除個別識別碼之普（抽）查資料，使用者於裝有監視設備與門禁管制之獨立空間進行作業，僅提供使用者不連網之電腦設備，資料處理場域禁止使用可攜式

設備或儲存媒體，且不得擅自以照相或其他方式，攜出電腦執行畫面及結果。

#### 四、遠端連線作業

刻正規劃新增之供應方式，遠端連線作業提供三大普查資料經去識別處理及移除較具敏感性資料，因本作業採遠端連線方式，管控機制仍有技術及實務上的限制未臻嚴密，為確保個資隱私安全，資料須經適當概化（generalization）

及抑制（suppression）處理。

本作業係於主計總處虛擬桌面基礎架構（VDI）下，透過專屬帳號及密碼，連線至獨立作業環境進行資料處理（附圖），申請單位須提供欲連線作業電腦之對外 IP 資訊，以建置專屬之 VDI 作業平台，藉由帳號驗證機制進行權限控管及軌跡稽核，以確保普查資料安全，檔案輸出入皆須透過具監控機制之隔離電腦傳輸至檔案伺服器，輸出資料僅限彙總之

統計結果，且須經審查後加密交付申請單位。

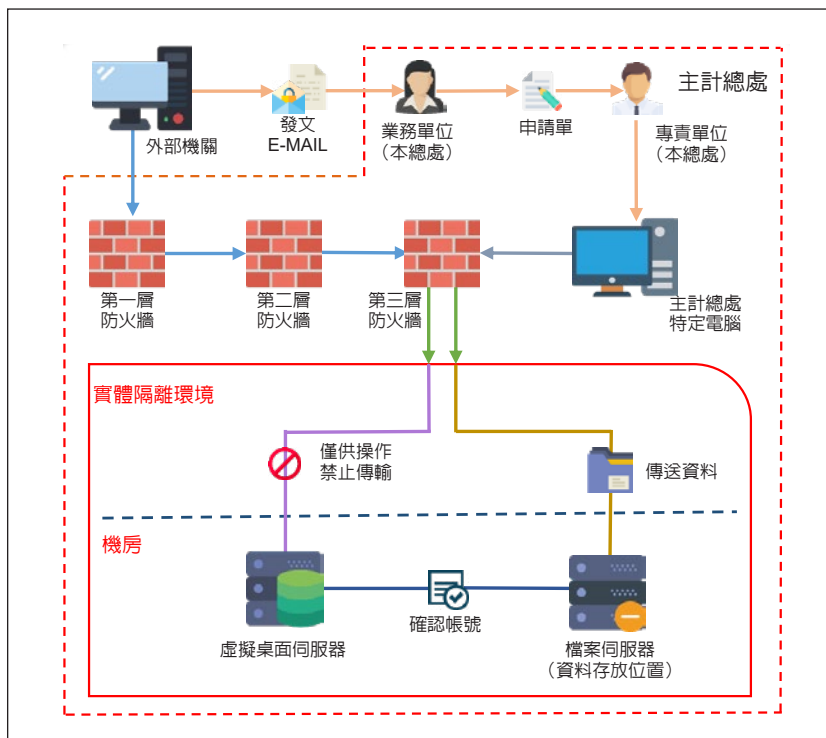
#### 肆、結語

調查資料的有效應用，是提升資料價值的體現，為順應作業環境及資料需用的變遷，總處參酌國際做法，重新檢視資料供應情形，研訂普（抽）查資料供應方式，依資料被揭露風險程度分類，並新增公開提供及遠端連線作業2種方式，冀能加強調查資料隱私保護，並擴大資料應用範疇及使用便利性，未來將持續關注相關發展，以優化調查資料供應作業。

#### 參考文獻

1. 黃怡婷、蘇南誠，張喻、林宥辰、陸薇安（2018），農林漁牧業普查外釋資料抽樣檔建置，行政院主計總處委外研究。
2. 美國普查局 <https://www.census.gov>.
3. 英國國家統計局人口普查微觀資料 <https://www.ons.gov.uk/census/2011census/2011censusdata/censusmicrodata>.
4. 澳洲統計局微觀數據應用 <https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Microdata+Entry+Page>.
5. 日本國家統計中心 <https://www.nstac.go.jp/services/archives.html>.
6. 韓國統計局微數據整合服務 <https://mdis.kostat.go.kr/index.do> ❖

附圖 遠端連線作業環境示意圖



資料來源：行政院主計總處國勢普查處。