

探析國際人口普查微數據之供應情形

為兼顧資訊安全及資料應用，本文彙集世界主要國家人口普查微數據對外供應情形，期供為我國做法之重要參考。

謝博行（行政院主計總處國勢普查處專員）

壹、前言

隨最新（西元 2020 年前後）世界人口普查結果陸續公布，各國普查資料均呈現豐富的人口與住戶資訊，供政府單位作為施政、資源配置及議席分配等用途。普查各項主題統計結果係經既定項目彙整數據（Aggregate data）編製而成，雖可讓一般人快速掌握各項人口及住戶之特徵資訊，惟較欠缺應用彈性；而微數據（Microdata）之外釋，適可彌補其不足處，資料使用者可依需求，深入探討各面向之主題

及分析。

由於美國與英國人口普查公眾微數據抽樣檔（Public Use Microdata Sample, PUMS）可經由網頁下載，取得容易，有助提高人口普查資料應用機會，另 IPUMS（Integrated Public Use Microdata Series）是由美國明尼蘇達大學、各國國家統計機構與國家檔案館等單位共同合作之國際人口數據庫，致力於收集、保存及傳播來自世界各地的人口普查與調查之公眾微數據，免費提供研究人員探討人口議題之所需。2000 年以來受惠於各國統計機構與歷史

普查資料之數位化，研究人員能取得的人口微數據越來越多（下頁圖 1），大幅增進國際人口資料應用之深度與廣度。本文將介紹美國、英國人口普查微數據提供情形，以及 IPUMS 數據庫蒐集與供應現況。

貳、國際人口普查微數據提供情形

各國普查當局基於普查資料安全考量，依不同提供方式規劃微數據產品。各類型微數據產品除直接刪除姓名、地址、出生日期等可識別項目外，尚須經過嚴格之去識別化處理作

業，因此不具有可辨別個人或家戶的資訊。茲就美國及英國做法簡介如次：

一、美國

美國普查局為補足使用者無法由公布統計結果中獲得之資訊，爰將普查資料經由抽樣及資料處理，提供人口普查公開微數據抽樣檔，以協助使用者自行產製所需的特定估計值和統計表。

美國 2010 年人口普查採用系統抽樣產製 10% 樣本作為 PUMS，可以直接透過網頁下載；2020 年人口普查更發展了

新的避免揭露系統（Disclosure Avoidance System），加強個別資料重新識別的保護，預計將以全新隱私保護做法產製之微數據產品（Privacy-Protected Microdata File）取代 2010 年人口普查 PUMS 資料，具體發布時間尚未確定。

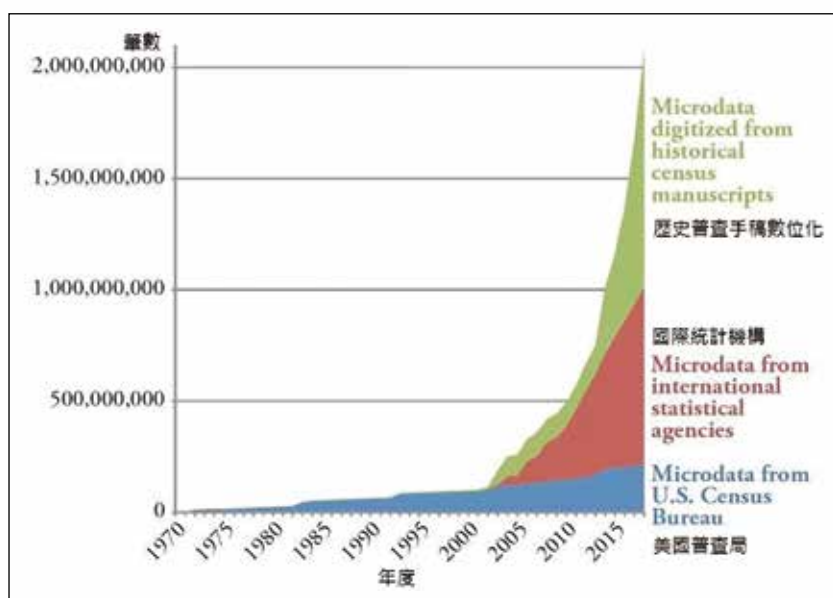
自 2010 年起，美國社區調查（American Community Survey）成為人口普查長表的替代資料，按年蒐集全國人口之社會、經濟、住宅和人口特徵等項目；該調查定期釋出 1 年期與 5 年期住宅樣本作為 PUMS，其中 1 年期 PUMS

包含全國 1% 樣本，而 5 年期 PUMS 則由 5 個 1 年期集結而成，並更新項目屬性與重新調整權重，以符合最新人口估計。美國普查局為便利民衆使用，爰於其人口普查網頁（<https://data.census.gov>）提供線上微數據分析工具（Microdata Access Tool），使用者無需其他軟體，僅選取所需項目可產製客製化表格與下載微數據。

為保護受訪者隱私，並且提供更精細的地理區域資訊，2021 年美國普查局提出以合成數據方式產製美國社區調查 PUMS 資料的想法，然而各界反應模擬資料不太適合用於創新研究。美國普查局刻正嘗試變通的做法，測試合成數據搭配驗證服務的可行性，意即當使用者使用合成數據進行分析後，可以將結果提交系統，藉由與真實數據比對，以評估數據有效性。

至美國人口普查與美國社區調查更小地區微數據屬管制資料，必須於美國聯邦統計研究數據中心（Federal Statistical Research Data Centers, FSRDC）臨場使用，FSRDC 係與學術機

圖 1 研究人員可取得的人口微數據筆數



資料來源：Steven Ruggles(2014), Big Microdata for Population Research, Demography 1 February 2014, 51(1), 287-297. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3949202>

專題

構和聯邦儲備銀行合作，專門處理機密微數據的安全環境，全國設有 33 個服務據點，以符合資格的研究人員就近使用。2022 年起，亦配合 2018 年循證決策基礎法案 (Foundations for Evidence-Based Policy making Act of 2018) 建立微數據標準申請流程，研究人員可於單一網頁窗口 (ResearchDataGov) 申請美國普查局與各聯邦政府統計機構之跨機關受限制微數據，以強化聯邦政府的實證構建能力。

二、英國

英國預計於今 (2023) 年第 2 季發布 2021 年人口普查微數據產品，大致延續 2011 年普查資料提供方式。微數據產品將依提供對象分為公眾 (Public)、受保護 (Safeguarded) 及安全 (Secure) 三種類型 (表 1)。其中公眾類型一般大眾可由網路自行取得；受保護類型為符合條件與同意服務條款之數據分析人員於英國數據服務網站 (UK Data Service) 註冊成功，才能取得微數據資料，主要服務對象為學術組織人員；安全類型則是專為經核准或認可之

研究人員準備，僅能在英國安全研究服務 (Secure Research Service) 之受管制環境使用。

英國公眾微數據之設置目的為滿足大眾獲取微數據的需求，推廣人口普查微數據的使用，並兌現政府開放資料的承諾，以推廣微數據之應用；其亦可作為「真實」數據的教育工具，協助統計和社會科學的教學，幫助學生了解微數據與已發布的統計表使用上的差異，稱為教學微數據 (Microdata Teaching File)。教學微數據產製方式係根據人口關鍵特徵將全國人口劃分為多個分組，由每個分組中隨機抽取 1% 個人樣本，使各特徵

比例能夠代表人口分配情形。教學微數據為區域層級資料，選擇地理範圍較小的地方行政區層級作為分組特徵，可以使樣本分布更均勻，避免出現極端樣本。

受保護微數據產製方式與公眾微數據相似，可提供更多變數細節與樣本數，並依地理區域大小分為區域 (Regional) 與地方行政區 (Local Authority) 兩種層級之 5% 個人樣本，地方行政區層級資料因地理區域分類較為細緻且人口較少，項目類別提供較為粗略。以年齡為例，區域層級資料可為單齡分組，地方行政區層級則為 5 歲年齡分組，以保護受訪者隱私。

表 1 英國 2021 年人口普查微數據產品

微數據產品	類型	抽出率	項目
公眾 (Public)			
Public-access sample (Microdata Teaching File)	個人	1%	20
IPUMS sample (2021 新增項目)	住戶	1%	30
受保護 (safeguarded)			
Safeguarded individual region sample	個人	5%	120
Safeguarded individual grouped local authority sample	個人	5%	120
Safeguarded household sample (2021 新增項目)	住戶	1%	50
安全 (secure)			
Secure individual sample	個人	10%	200
Secure household sample	住戶	10%	200

資料來源：Office for National Statistics, <https://www.ons.gov.uk>

安全微數據提供個人樣本與住戶樣本兩種產品，個人樣本包含 10% 居住在住戶與公共機構（如宿舍、醫院等）的人口；住戶樣本包含 10% 住戶及戶內人口資料，兩種樣本不互重疊，不能合併使用。

英國根據過去使用者之反饋，2021 年人口普查微數據將新增兩項產品，第一項產品係於受保護微數據新增區域層級之 1% 住戶樣本，約 50 個項目，由於個人資料加入家庭資訊，將提高個人身分辨別的機會，住戶樣本規劃僅提供某些家戶特徵，且同居伴侶者必須與結婚者合併，同性伴侶不被揭露，以及評估職業改以行業及社會經濟分類替代（The National Statistics Socio-economic classification）；第二項產品為配合美國明尼蘇達大學 IPUMS 收集各國人口微數據計畫，提供以區域層級之 1% 住戶樣本，項目相對較少，約 30 個項目。

三、IPUMS 數據庫

美國自 1962 年開始發布人口普查微數據，而其他對於微數據提供之進展非常緩慢，

大部分國家在發布彙總統計結果後，並沒有保存或再利用微數據之機制，因此研究人員可用的微數據數量非常有限，更難以進行跨時間或跨國家人口研究比較。迄 1990 年初期，IPUMS 之美國項目開始發展，為美國人口普查建立格式一致的微數據資料，並導入複合編碼方案處理變數編碼，例如：與戶長關係項目，前兩碼為歷次普查資料中可共通的類別，加上後兩碼為僅在個別普查年度中出現的細項類別（表 2），可使每次普查之主要類別編碼不會隨時間改變，同時保留各次普查類別間差異資訊，資料更易於跨時間比較，節省許多研究人員的整理資料時間。1999 年獲得美國國家科學基

金會和美國國立衛生研究院的資助，接續推動 IPUMS 國際項目，增加收集美國以外其他國家人口微數據，目前約收集 103 個國家，547 次人口普查和調查資料（下頁圖 2）。

IPUMS 為微數據傳播建立線上數據抽取系統，有別於一般提供單個檔案方式，研究人員可依需求自訂選擇多個國家、資料年度、家戶與個人項目及樣本比例等條件，系統將會自動產生請求的資料，並可線上分析或下載檔案，研究人員無須處理無相關的資料或承擔額外的風險，使資料使用更加便利與安全。IPUMS 資料來源主要是依賴夥伴國家的贊助，資料提供取決於夥伴國家處理微數據時間與後續 IPUMS

表 2 IPUMS 複合編碼對照示例（與戶長關係）

編碼 (Code)	IPUMS 編碼對照
0101	戶長
0201	配偶
0202	第二 / 第三任妻子（一夫多妻制）
0301	子女
0302	收養的子女
0303	繼子女

資料來源：IPUMS USA, <https://usa.ipums.org>

專題

整合資料時間，因此 IPUMS 釋出微數據的時點較難掌握，2022 年發布新增智利（2017 年）等 7 個人口普查樣本。

近年 IPUMS 為儘可能提供使用者詳細的地理資訊，已開始對數據充足且滿足機密性的樣本中新增小地區地理變數，並附帶製作地圖所需的邊界檔案（Shapefile），期結合地理變數使用，繪製更詳細的人口趨勢圖；2022 年已釋出 20 個具有三級行政級別地理數據樣本。小地區地理變數之產製雖面對許多挑戰，如來源資料缺乏細節地理資訊、地理變數需要花費大量時間與真實之

邊界標籤比對，並可能出現沒有足夠的地圖或邊界檔案等問題，IPUMS 仍持續增加更多國家小地區地理變數，以滿足資料應用需求。

參、結語

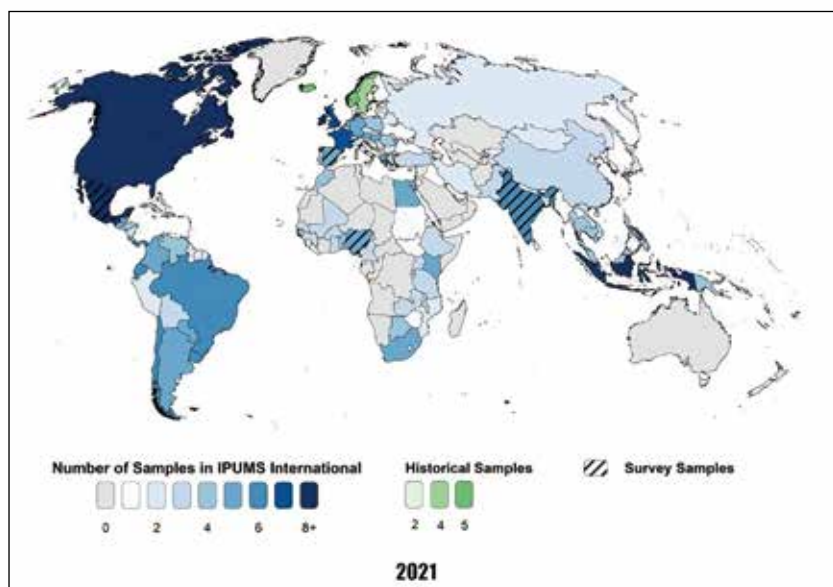
人口普查微數據之提供必須同時兼顧資料隱私保護與使用性，PUMS 提供使用微數據之便利性，給予的民眾參與研究人口普查資料的機會，使用者扮演之角色由過往之被動獲取資訊轉為主動探索，實有助於推廣人口普查之使用；微數據之內容提供，更應多方蒐集使用者經驗，資料編碼、資

料格式以及檔案讀取方式等，才能開發更貼近使用需求的產品，降低使用者入門難度。近年，英美除了改進微數據產品項目，亦積極優化管制環境功能，建立微數據整合平台，提高管制環境可連結的微數據數量，大幅增加微數據實用性；而 IPUMS 所收集的人口微數據數量愈趨增加，更有利於國際人口議題研究，相關做法適可供為我國普查資料供應之重要參考。

參考文獻

1. Census Bureau, <https://www.census.gov>
2. Changes To Census Bureau Data Products, IPUMS, <https://www.ipums.org/changes-to-census-bureau-data-products>
3. Donna Daily (2022) Disclosure Avoidance Protections for the American Community Survey, American Community Survey Office.
4. IPUMS INTERNATIONAL, <https://international.ipums.org>
5. Kugler, T. and Fitch, C. (2018), Interoperable and accessible census and survey data from IPUMS. Scientific Data volume 5, 180007 (2018).
6. Office for National Statistics, <https://www.ons.gov.uk>
7. Quinn Heimann (2022), IPUMS International has brand new low level geographic variables and shapefiles, <https://blog.popdata.org> ❖

圖 2 IPUMS 國際項目各國樣本個數



資料來源：IPUMS International 2022 Data Release, IPUMS Blog, <https://blog.popdata.org/ipums-i-2022-data-release/>