

美國各類母體資料庫建置及應用情形

美國商務部普查局等機關所建置之母體資料庫已行之有年，經驗及技術均甚為完備。本文簡介其各類母體資料庫發展情形、系統架構及應用概況，作為各單位精進相關作業之參考。

◎ 吳馨和 (行政院主計處第4局專員)

壹、前言

我國工商及服務業普查因每5年辦理1次，辦理週期較長，非普查年缺乏當期母體資料供各界應用，行政院主計處爰於99年3月起推動「工商及服務業母體資料檔建置計畫」，藉相關資料之蒐集彙整，建置並常川更新母體資料檔，俾獲致當期產業發展資訊，掌握母體資料時效。惟母體資料檔建置作業浩繁，而美國對於建置各類母體資料庫行之有年，技術甚為先進完備，

應用層面亦廣，爰針對其相關業務內涵概括介紹，藉參考國外成功案例，以提升工商及服務業母體資料運用效益。

貳、美國各類母體資料庫建置及運作情形

因美國各類母體資料之應用層面甚廣，爰依其用途，責由各機關分別建置業管之母體資料庫，其中與我國工商業母體資料檔特質與屬性最相近者計有三項，分別為商務部普查局 (Census Bureau) 工商業

母體資料庫 (The Business Register, BR)、勞工部勞工統計局 (Bureau of Labor Statistics, BLS) 母體資料庫 (The Business Establishment List, BEL) 及農業部國家農業統計署 (National Agricultural Statistics Service, NASS) 母體資料庫 (The Farm Register, FR)，爰針對其建置法源、資料項目、系統架構及應用情形分述如次：

一、普查局工商業母體資料庫 (BR)

(一) 法源依據：BR係依據美國法律（The United States Code，U.S.C.）標題13普查法（Titles 13，Census）、標題26國內稅收法（Titles 26，Internal Revenue Code）及國稅局準則第26項（26 CFR 6103）建置、維護及應用。其中普查法第1章即明文規範：商務部部長只要認為統計業務確有需要，可要求任何其他聯邦政府部門、機構及哥倫比亞特區政府部門，提供相關資料。第10章亦規範：普查局為統計目的，根據書面協議，在機密資訊保護法規定下，可提供資料予經濟分析局及勞工統計局或指定的統計機構，以符合統計效率。

(二) 資料項目：BR係以廠商設立（Establishment-Based）資料為基礎，並

加入經濟普查及相關抽樣調查資料，各資料欄位中，以企業代號（Employer Identification Number，EIN）及個人社會安全號碼（Social Security Number，SSN）為惟一鍵值，二者均屬企業納稅實體（Taxpaying Entity）之代號（詳表1）。

(三) 系統架構：BR係以甲骨文資料庫（Oracle Database）為系統建構主軸，且以關聯（Related Table）方式連結相關資料庫，並藉線上互動式介面（Interactive Web-Base Interface）及PL/SQL（Procedural Language/SQL）呈現及查詢資料。BR因資料量龐大，資料庫系統功能模組應用程式皆須有極高之效能。BR系統架構以抽樣調查及普查資料庫、EIN資料庫、SSN

資料庫、雇主資料庫、地址及地理編碼資料庫及MIS人員資料庫為主，並以EIN及SSN連結各項表格資料，以更新資料庫內容。甲骨文資料庫系統雖價格昂貴、維護成本及技術需求均甚高，惟因具備高延展性（可執行於各種系統平台）、支援所有資料（各種型態資料）、資料庫可靠性（具備容錯能力）、高效能（Performance）及開發工具（Development Tools）完備等多項優勢，爰為BR系統架構之首選。

(四) 應用情形：BR可於相關法規規範下，供為各機關廠商面抽樣調查之母體；並以維護廠商隱私資料為前提，提供國家整體經濟研究及分析之資料來源，俾了解當期廠商營運情形。

表 1 BR項目內容

項目內涵分類	項目內容	說明
識別個別企業	企業代號 (EIN)	企業持續經營 (Permanent) 則其代號有效且惟一 (Unique)
	個人社會安全號碼 (SSN)	
通信調查	企業名稱	
	通訊地址	
	電話	
	傳真	
分類、範圍統計	電子郵件信箱	
	業別代號 (NAICS)	NAICS指北美行業分類系統 (North American Industry Classification System)
	地理代號	地理資訊系統 (GIS) 代號
	報稅狀況	
企業經營活動、母體分層抽樣變數、研究企業經營特徵 (CBP, County Business Patterns) 及產製相關統計表	廠商類別	單一場所企業 (SU)、多場所企業 (MU)
	員工薪資	
	企業銷售數量	
	企業單據	
	營業收入	
	營業費用	
	期初存貨	
期末存貨		
資料處理	經營狀態	含營業中、歇業、倒閉、出售、其他
	區域代號	含州、郡、市鎮、普查區、街廓
	抽樣狀態	是否抽樣
	樣本加權資料	樣本加權、擴大數
	調查表式代號	識別不同調查表式，掃描BAR CODE
	檢查代號及通信代號	識別通信調查表
	前端編輯代號	資料編輯異動代號

二、BLS母體資料庫 (BEL)

(一) 法源依據：BEL係依據美國失業保險法 (Unemployment Insu-

rance laws)，由BLS蒐集全美各州之勞動資料 (ES-202表格)，彙整並建置資料庫。

(二) 資料項目：BEL涵括全美1億3,700萬全職

(Full Time) 及兼職 (Part Time) 員工所屬廠商資料，可彙總6兆4,000萬美元薪資及910萬場所單位家數資訊 (項目詳表2)，惟其範

圍不含農業耕種者、非營利組織、宗教機構學校、學生、被拘留者、自營作業者（Self Employed）及無酬家屬工作者（Unpaid Family

Workers）等。

（三）系統架構：BEL系統架構係甲骨文及Sybase資料庫，屬中大型資料庫，其價格、成本、技術需求及特點均與BR相

同，惟其資料量略少於BR。

（四）應用情形：BEL建置之初始目的為掌握失業保險員工人數，迄今其應用範圍則擴大至各勞動

表2 BEL項目內容

資料名稱	項目內容	備註
雇主主檔 (Employer Master File)	企業代號 (EIN)	雇主識別碼
	單位級別	區分企業或場所
	通訊地址	
	實際地址	
	統計地址	
	企業名稱	
	開業日期	
	開始負責日期 (Initial Date of Liability)	
	終止負責日期 (End of Liability Date)	
	復業日期 (Reactivation Date)	
	失業保險代號 (Unemployment Insurance (UI) Code)	同一州 (State) 的UI惟一，但跨州則不惟一
	企業員工人數	
	場所員工人數	Report Unit Number (RUN)
	聯邦雇主識別代號	由國稅局 (IRS) 分配雇主代號 (繳稅用、確定雇主稅率及BLS連結各州雇主用途)
企業長期編號 (Longitudinal Database Number)	連結同一企業不同年度資料	
組織別	業主 (Proprietor)、合夥、公司及其他	
經營權別	民營、公營 (聯邦、州及市鎮)	
會計檔案 (Accounting File)	員工總薪資	最近2年資料
	員工含稅薪資	
薪資檔案 (Wage Record)	社會安全號碼 (SSN)	
	企業主名稱	
損益歷史檔 (Benefit History)	企業損益	最近2年資料
宣告檔 (Claimant File)	企業主社會安全號碼	
	企業損益	

主題之抽樣調查母體及勞動經濟研究分析，供國民所得統計、經濟預測、地方經濟發展，以及勞動市場之就業、失業、轉業等相關分析應用，其縱貫資料 (Longitudinal DataBase) 亦可用以分析產業長期發展趨勢，應用價值極高，爰為政府釐定施政計畫之重要資訊來源。

三、NASS母體資料庫 (FR)

(一) 法源依據：本資料庫係依美國法律 (The United States Code, U.S.C.) 標題7農業法 2276章 (Titles 7, Agriculture, Section 2276)、機密資料保護法及統計效率法 (Confidential Information Protection and Statistical Efficiency Act) 建置，且NASS須針對統計調查取得之資料負保

管責任。

(二) 資料項目：本資料庫內容為農業經營者與從事農業企業 (Agriculture Business) 之資本主及其農地面積、牲畜數等相關資料 (項目詳表3)。

(三) 系統架構：自1992年起，NASS決定以Sybase資料庫取代原RTMMS系統 (Real Time Mail Maintenance System)，作為母體資料庫ELMO系統 (Enhanced List Maintenance Operations) 建置主軸，考量重點

如次：

- 1.效能佳、錯誤率低、執行速度快及系統穩定性高。
- 2.執行系統之硬體需求低，硬體維護成本低，系統記憶體需求低。
- 3.資料庫管理容易，專業人員養成時間較短，人力成本低。
- 4.Sybase除以ODBC (Open DataBase Connectivity) 資料庫界面存取資料外，亦可使用原生驅動程式 (Native Driver) 連接資料庫，直接與資料庫前端驅動程式連接，大幅提升

表3 FR項目內容

變數型態	項目內容	備註
識別變數	姓名	
	通訊地址	
	電話	
	區域代號	含州、郡、市鎮
	社會安全號碼 (SSN)	
	企業代號 (EIN)	
控制變數	土地總面積	英畝
	個別農作物面積	英畝
	糧食倉庫容量	
	最多牲畜存貨	
	最多僱用員工人數	
統計變數	每年彙總結果	
	數量	

資料存取速度。

5. 易於移轉資料 (Migrate)

，可將重要資料移至成本較低的高速平台，且同時保護舊有系統。

6. 開發工具需求低，操作介面簡易，採用模組化及彈性化之設計，縮短程式開發時間。

7. 支援Java 及以Java 建立元件化的資料庫應用程式，及無線化 (Unwire) 系統架構。

(四) 應用情形：ELMO 系統包含抽樣子系統 (Generalize Enhanced Sampling and Information System，簡稱 GENESIS)，其包含分析工具 (Profiler)、母體定義 (Population Definition)、抽樣 (Sampling) 及分析 (Analysis) 等4 個功能模組 (Module)，可即時產生相關抽樣調查樣本資料，抽樣功能完備；並提供相關研究分

析應用，供農業部及相關農政機關參考。

四、各母體檔更新作業及系統架構

依據美國普查法，普查局除本身機關內的普查及抽樣調查資料外，只要商務部長認為確有需要，可要求相關聯邦政府部門，提供各機關內之抽樣調查及相關公務檔案資料，據以常川更新BR。另為使BR 內的「業別代號 (NAICS)」維持最新資料，每季由普查局挑出EIN 鍵值 (僅EIN 鍵值無原始資料)，交由BLS 比對業別代號，經比對成功 (約50% ~ 55%) 並確認後，再回覆具業別代號之資料，供BR 更新作業，以獲得即時及正確的業別代號。

至於BEL 及FR 二項資料庫之更新作業則係以主管機關內部資料為主要來源。BLS 以機關內之抽樣調查資料、公務登記資料及公務登記個別資料常川更新BEL；NASS 除抽樣調查及公務登記資料外，尚包

含農業普查資料，亦屬FR 之更新來源之一。茲比較前揭3 項母體資料庫建置作業及系統架構如表4 所示。

參、結語

綜上所述，美國母體資料庫建置作業，無論就個別資料保護、建置法源及相關應用技術而言，頗多值得借鏡之處；而我國於民國99 年起推動建置並常川更新母體資料，時間尚短，仍需持續精進。爰參酌美國作法，列述未來母體資料建置之發展方向如次：

一、明確制訂母體資料建置法

源：因母體資料建置須具備充分更新來源，為避免隱私外洩情事發生，美國各母體更新資料之業管機關皆依據相關法律規定，於維護隱私之原則下，配合建置作業提供所需資料。為永續發展本項重要業務，我國相關法規實宜仿效美國作法，將母體更新資料提供、母體應用及

隱私維護等事項，於相關法規（如統計法）中明文規範，俾藉法律位階及規範明確性之雙重提升，降低作業執行困難。

二、精進母體資料建置作業技術層次及應用便捷性：依據我國「工商及服務業母體資料檔建置計畫」作業內容，母體資料係以檔案型式，於IBM大型主機MVS（Multiple Virtual Storage）系統建置，並以SAS、FORTRAN及COBOL等軟體開發程式，執行速度較慢，且部

分資料需經人工確認，建置初期爰以資料常川更新為主要目的。未來宜充實軟、硬體與人力需求，朝美國資料庫建置模式發展，並將抽樣作業、資料擷取及連結相關資料作業融入系統模組，俾臻整體系統之一致性及完備性，並在確保隱私維護原則下，提供即時、正確而便捷之母體資訊供各界應用。

三、擴大母體資訊應用層面：美國各項抽樣調查母體來源皆為所屬之母體資料

庫，因母體來源相同，更新速度及範圍一致，調查品質爰隨之提升。反觀國內廠商面抽樣調查，多應用辦理機關之業管登記資料為母體，致發生母體範圍、更新時間及資料項目不一致或不完整等問題，而影響調查資料品質。未來我國除應針對個別資料性質，由相關機關建置不同之母體資料外，亦應鼓勵或輔助調查辦理機關，將抽樣母體齊一化，俾精進調查品質，並擴大母體資訊應用層面。❖

表4 美國母體資料庫建置作業及系統架構比較表

	BR	BEL	FR
建置法源	有	有	有
辦理機關	普查局 (Census Bureau)	勞工統計局 (BLS)	國家農業統計署 (NASS)
資料庫系統	Oracle	Oracle & Sybase	Sybase
作業平台	Windows & Linux	Windows & Unix	Windows & Linux
建置時期	1968年	1990年	1992年重新開發
效能	佳	佳	佳
資料型態	大型資料庫	中大型資料庫	中型資料庫
專業人力技術	高	高	較低
維護成本價格	昂貴	昂貴	較低
開發工具	靈活且具彈性	靈活且具彈性	簡潔具親和力
特性	支援建立元件化的資料庫應用程式	支援建立元件化的資料庫應用程式	容易移轉資料 (Migrate)
應用情形	抽樣母體及工商服務業相關分析	抽樣母體及勞動市場、勞動條件相關分析	抽樣母體及農業相關分析