

編號：99 年 03

行政院主計處研究報告

普查母體建置採用資料庫處理之研究 —以人口及住宅普查為例

電子中心分析設計組

江桂英、謝明憲

行政院主計處

印製日期：100 年 3 月 4 日

摘要

行政院主計處現行辦理之基本國勢調查包含「人口及住宅普查」、「農林漁牧業普查」及「工商及服務業普查」三大普查。在調查進行前，必須先建置母體資料，根據母體資料進行普查區劃分、普查區名冊列印等前置作業，再根據母體資料中的地址前往訪查，因此母體建置作業是普查的根本。而目前三大普查沿襲過去經驗於 IBM 大型主機環境進行母體建置相關資料處理，惟普查母體資料量大，建置繁瑣，加以主機平台人才日漸缺乏且介面較不友善，因而產生採用資料庫建置普查母體之需求。

因此，本研究採用資料庫進行人口普查建置之可行性研究與實做，希望增進查詢親和性，減少程式開發負擔，簡化公務檔案處理流程以提升效率且減少人力耗費，並進而利用資料庫功能進行母體疑義資料處理，提昇母體資料品質。

而因人口及住宅普查母體規模為三大普查之最，且主要資料處理作業由行政院主計處電子處理中心負責，加以戶役政資訊系統門牌資料為全國最完整的地址官方登記資料，建置成果可為農林漁牧業普查及工商及服務業普查應用，故本研究以人口普查母體建置之資料處理作業為範圍，探討採用資料庫進行處理之可行性與效益。

本研究首先確認研究目標與範圍，接著搜集相關資料，經歸納整理後建構出本研究之研究架構。然後針對本研究之範圍進行需求分析與設計，並實際採用資料庫建置人口普查母體，並與現行母體進行勾稽比對，最後做出結論與建議。

其中，本研究發現採用資料進行母體建置，其介面操作較為友善，且資料庫結構化查詢語言可減少應用程式開發負擔。而公務檔案處理流程可搭配資料庫網要設計進行簡化，除可降低處理複雜度，並可保留原始地址欄位及調整後欄位之對應，於後續母體資料維護作業僅處理異動資料，減少工作負擔。而本研究所建置之母體與現行母體資料勾稽後發現勾稽成功率高達 99.92%，母體資料品質符

合需求。另，可使用資料庫建置之人口普查母體資料為基底，建置模糊比對功能供其他調查連結地址，以降低各調查為提高連結率所耗費之時間及人力成本。

最後，由於本研究結果顯示採用資料庫進行母體建置具有親和性及效率上的優點，故建議擇定工商及服務業普查作業建立雛型系統，導入資料庫處理作業模式，進而透過教育訓練，推廣至三大普查母體建置。而工商及服務業普查母體建置亦需配合規劃相關規劃資通安全機制，包含權限管理、存取稽核紀錄及營運持續性計畫等，始能保障個人資料與組織業務運作安全。而母體資料庫建置完成後，亦可據以整體規劃普查商業智慧環境，除名冊列印等前置作業外，整合實際訪查作業後續之調查資料處理、公務檔案連結、調查資料檢誤、分析以及結果表編製與列印進行整合規劃，導入萃取轉製載入（ETL）工具支援公務檔案處理，結合既有之光學字元識別（OCR）自動化調查資料處理，且整合資料倉儲（Data Warehousing）、線上分析處理（OLAP）及報表編製軟體，使普查價值進一步提升。

目 錄

摘 要.....	一
目 錄.....	三
圖 目 錄.....	四
表 目 錄.....	1
第一章 緒論.....	2
第一節 研究動機與目的.....	2
第二節 研究範圍.....	3
第三節 研究流程.....	3
第二章 普查母體建置現況分析與探討.....	5
第一節 三大普查母體建置現況分析.....	5
第二節 人口普查母體建置資料處理需求及流程概述.....	7
第三節 人口普查母體資料來源及交換格式.....	1
第四節 現行資料處理環境說明.....	9
第五節 現行中文轉碼及罕見字處理程序.....	9
第六節 現行地址階層化及疑義地址處理程序.....	10
第三章 研究過程及成果.....	15
第一節 資料處理環境與範圍說明.....	15
第二節 中文轉碼及罕見字處理流程.....	16
第三節 地址階層化處理流程.....	19
第四節 疑義地址處理流程.....	23
第五節 人口普查母體資料庫品質.....	28
第六節 執行時間及人工參與分析.....	30
第七節 外國人口母體建置—地址連結應用.....	31
第四章 結論與建議.....	35
第一節 結論.....	35
第二節 建議.....	37
參考文獻.....	39

圖 目 錄

圖 1-1 研究流程圖	4
圖 2-1 套印調查表	8
圖 2-2 普查名冊封面	9
圖 2-3 普查名冊內頁	10
圖 2-4 普查名冊內頁宅、戶、人資料	10
圖 2-5 普查區地圖	11
圖 2-6 街道範圍一覽表	12
圖 2-7 人口普查母體建置資料處理流程示意圖	1
圖 2-8 現行人口普查母體建置資料處理環境示意圖	9
圖 2-9 普查名冊列印順序示意圖	11
圖 2-10 戶役政資訊系統村里門牌資料地址欄位格式	13
圖 3-1 本研究資料處理環境示意圖	15
圖 3-2 java.util.regex 「號」階層範例	21
圖 3-3 戶役政資訊系統村里門牌地址階層化資料庫畫面	22
圖 3-4 本研究母體資料名冊資料庫畫面	29
圖 3-5 門牌地址異動情形	31
圖 3-5 內政部入出國及移民署外僑人口檔畫面	32
圖 3-6 字串模糊比對推薦連結地址	34
圖 4-1 以資料庫查詢語言查詢臺北縣路街道名稱最長者	35
圖 4-2 新增或異動地址比對	36

表 目 錄

表 2-1 人口普查應用表件.....	13
表 2-2 人口普查母體資料來源及交換格式一覽表.....	1
表 2-3 現行地址階層標準格式.....	11
表 2-4 現行地址欄位階層格式.....	13
表 3-1 常見中文編碼.....	16
表 3-2 EUC 與 CNS 碼對應規則.....	17
表 3-3 地址常見罕見字.....	18
表 3-4 無人設籍門牌地址比率.....	19
表 3-5 簡併及調整地址欄位階層格式原則.....	19
表 3-6 簡併及調整後地址欄位階層格式.....	20
表 3-7 疑義地址過濾方式.....	23
表 3-8 關鍵字順序、格式或次數錯誤地址人工處理.....	23
表 3-10 含「社區名稱」階層之地址.....	24
表 3-11 「地區名稱」階層含其他階層關鍵字之地址.....	25
表 3-12 「地區名稱」字串長度分布.....	25
表 3-13 「地區名稱」字串長度大於 5 之地址.....	26
表 3-14 「路街道」字串長度分布.....	26
表 3-15 「路街道」字串長度大於 6 或小於 2 之地址.....	27
表 3-16 本研究母體資料庫與現行人口普查母體資料勾稽結果.....	28
表 3-17 本研究與現行母體階層規則不同類型.....	28
表 3-18 簡寫及錯別字對照表.....	29
表 3-19 執行時間與人工參與分析表.....	30

第一章 緒論

第一節 研究動機與目的

行政院主計處（以下稱主計處）現行辦理之基本國勢調查包含「人口及住宅普查」（以下稱人口普查）、「農林漁牧業普查」及「工商及服務業普查」三大普查。由於普查（census）是對整個母體進行全面調查，而在全面調查進行之前，必須先建置母體資料，再根據母體資料進行普查區劃分、普查區名冊列印等前置作業，最後再根據母體資料中的地址前往訪查。因此，母體建置作業是普查的根本。

目前三大普查皆沿襲過去經驗於 IBM 大型主機環境進行母體建置及相關資料處理，母體資料以循序文字檔的形式儲存。業務單位根據各項資料處理需求於大型主機平台開發 Fortran、SAS 及 COBOL 等應用程式進行資料處理。惟普查之母體資料量大，建置繁瑣，加以主機平台雖穩定安全，但人才日漸缺乏，且介面較不友善，因而產生採用資料庫建置普查母體之需求。

以人口普查母體建置資料處理作業為例，人口普查母體建置作業使用內政部戶役政資訊系統之「村里門牌資料」及「個人戶籍資料」，目前我國約有 900 萬筆門牌資料及 2,300 萬筆戶口資料，而因建物及人口狀況隨時間不斷更迭，每次正式或試驗性調查進行前皆須重新取得資料進行處理，因而帶來巨大的資料處理工作量。此外，為滿足人口普查名冊須按「段」、「巷」、「弄」、「街」、「號」、「樓」順序排序之需求，地址欄位須切割階層儲存。然門牌及戶籍資料地址欄卻未區分階層，且關鍵字及格式繁多，導致地址階層化問題複雜，不易進行。且門牌及戶籍資料包含許多罕見字、簡寫字及因筆誤而不實際存在的疑義地址，皆須預先處理及排除，而門牌及戶籍資料量大，檢視及排除這些地址所需的人工作業量極大。且大型主機環境欠缺友善的資料查詢介面，而文字檔案之管理亦較不結構化，許多資料處理作業亦皆仰賴程式撰寫始能達成，形成本中心人力負荷。因此，本研究採用資料庫進行人口普查建置之可行性研究與實做，希望能達成下列目

標：

1. 採用資料庫建置普查母體資料，增進查詢親和性並減少程式開發負擔。
2. 簡化公務檔案處理流程，提升效率並減少人力耗費。
3. 採用資料庫進行疑義地址處理，兼顧母體資料品質與效率。

第二節 研究範圍

主計處三大普查包括人口及住宅普查（以下稱人口普查）、農林漁牧業普查及工商及服務業普查，其中，本中心支援人口及住宅普查母體資料處理作業。因人口及住宅普查母體規模為三大普查之最，且主要資料處理作業由本中心負責，加以戶役政資訊系統門牌資料為全國最完整的地址官方登記資料，其建置成果可為農林漁牧業普查及工商及服務業普查判定疑義地址或行政區域欄位闕漏時使用，故本研究以人口普查母體建置之資料處理作業為範圍，探討採用資料庫進行處理之可行性與效益。

第三節 研究流程

在研究流程方面，本研究先確認研究目標與範圍，接著搜集相關資料，經歸納整理後建構出本研究之研究架構。然後針對本研究之範圍進行需求分析與設計，並實際採用資料庫建置人口普查母體，並與現行母體進行勾稽比對，最後做出結論與建議。詳細研究流程如圖 1-1 所示。

1. 確立研究動機與目的

由我國基本國勢調查母體建置資料處理作業之環境與困難確立本研究動機、目的及研究範圍。

2. 現況分析與探討

針對現行人口普查母體建置資料處理作業之需求進行分析，據以採用資料庫

進行人口普查母體建置。

3. 採用資料庫建置人口普查母體

以現況分析與探討之結果搭配資料庫系統之建立來實做母體建置資料處理作業，並使用資料庫功能設計相關流程以增進母體品質。

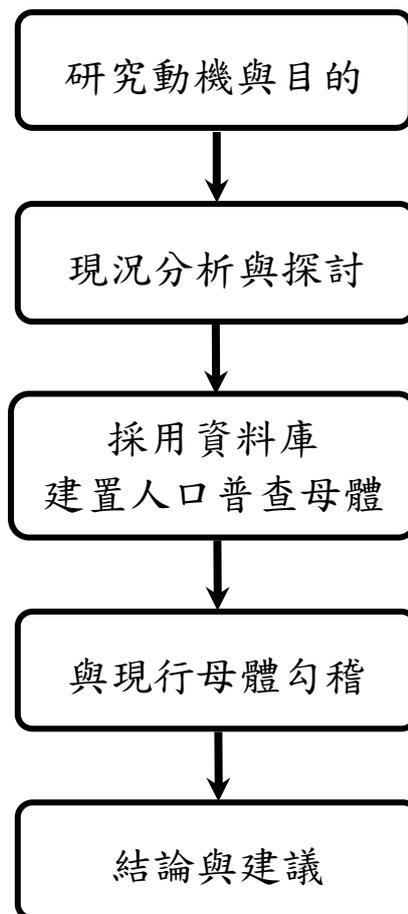
4. 與現行母體勾稽

將本研究建置之人口普查母體資料庫與現行母體資料進行勾稽比對，檢視品質是否符合需求。

5. 結論與建議

依據本研究實際建置過程之發現與分析，提出結論與相關建議。

圖 1-1 研究流程圖



第二章 普查母體建置現況分析與探討

第一節 三大普查母體建置現況分析

1. 人口及住宅普查：

- (1) 人口普查母體包含本國人口及外國人口部分，本國人口母體資料來源為內政部戶役政資訊系統之「村里門牌資料」及「個人戶籍資料」，外國人口母體來源為內政部入出國移民署提供之「外僑人口檔」。
- (2) 本國人口母體建置先針對村里門牌資料進行資料處理，續以門牌地址連結個人戶籍資料戶籍地址以確認每個門牌地址(宅)之設籍戶數與各戶籍所含人數，再將此資料匯入主計處人口普查 GIS 系統，以每個普查區包含 110 戶為原則，搭配自然地理界線劃分普查區，最後再根據劃分結果產製調查名冊及街道範圍一覽表，並列印地圖。外國人口母體建置則遵循本國人口普查區之劃分，以外僑人口檔「居留地址」欄位為鍵值與本國人口母體資料進行連結，確定普查區歸屬，並據以列印名冊。

2. 農林漁牧業普查：

- (1) 以最近一次農林漁牧業普查使用的母體資料為基底，每年利用農業專案調查抽樣的 20% 樣本回饋於農牧戶母體資料，並彙整各主管機關最新農林漁牧業經營者名冊如農保檔、全國土地財產檔、農戶耕地資料檔、農業產銷班成員檔、農漁牧業生產合作社社員檔、養雞協會會員檔、休閒觀光農場檔、退輔會農林漁牧場檔、動力漁船漁筏檔及無動力漁筏舢舨檔、養殖漁塭漁戶檔等，加以整理建置為下次普查母體資料。
- (2) 由於上述資料來源彙整過程須判別資料是否重複，故必須進行比對，比對鍵值先使用身分證字號判斷是否為同一戶，續使用地址確認該戶所屬行政區域，並統計各村里所含戶數資訊，以利基層調查行政單位分配普查員工

作負擔；惟中文地址於比對前需先經一致化處理，始能提高連結率。另，亦利用內政部戶役政資訊系統之村里門牌異動資料更新行政區重劃之普查母體資料。

3. 工商及服務業普查：

- (1) 以最近一次工商及服務業普查使用的母體資料為基底，加上稅籍登記資料、公司登記資料及勞保資料等相關公務檔案進行建置。後續並利用工廠校正調查及受僱員工薪資調查等抽樣調查回饋資料進行比對，以研判各類資料之正確性，並剔除倒閉停歇業廠商、新增新開業廠商及進行各種異動狀況更新，以更新母體資料。
- (2) 同農林漁牧業普查，母體各資料來源彙整過程必須判別是否重複，故必須進行比對，以營利事業統一編號為比對鍵值來判斷是否為相同訪查對象，續使用地址確認該戶所屬行政區域，惟中文地址於比對前需先經一致化處理，始能提高連結率。

第二節 人口普查母體建置資料處理需求及流程概述

現行主計處人口及住宅普查採派員面訪調查法及連結公務檔案產生普查資訊。為利調查作業順利進行，需適切規劃單一普查區所含戶數，以平衡普查員工作負擔；並須提供普查名冊、街道範圍一覽表及套印調查表等各式踏查與實地訪查用表件，以輔助普查員實地訪查作業的進行。而普查區劃分及各式踏查與實地訪查用表件的產製皆奠基於普查母體的建置，故於正式普查作業進行前便須建置普查母體資料，據以規劃後續相關作業。

現行人口普查母體資料來源為內政部戶役政資訊系統村里門牌資料及個人戶籍資料，母體建置作業所包含的資料處理程序包含中文轉碼、罕見字處理、疑義地址排除、地址階層化等資料處理，將門牌與戶籍資料連結以建立人口普查母體，進而據以產製普查名冊、街道範圍一覽表及套印調查表等表件。而人口普查GIS系統亦匯入普查母體資料，利用各門牌地址（宅）設籍戶數及每戶所含人數等資訊劃分普查區。

以下分別針對各踏查與實地訪查用表件說明用途及相關資料處理需求：

1. 套印普查表：

為輔助普查員實地訪查，本處特別針對各設籍門牌地址（宅）內的每戶，於調查表上套印戶內人口之姓名、性別、出生年月日及身分證號後 5 碼等個人資料，以提升普查員尋及受訪目標進行相關調查事宜之效率；套印調查表格式圖 2-1。

2. 普查名冊：

為輔助普查員實地訪查及相關行政管理作業，主計處列印各普查區名冊，內容包含該普查區所屬之宅（門牌資料）、戶與人（戶籍資料）資料，供普查員實地訪查參考，並提供訪查狀態註記欄位以利普查行政管理作業進行。普查名冊樣式如圖 2-2 及圖 2-3。

圖 2-2 普查名冊封面

99 年人口及住宅普查名冊								
普查員用								
普查實施期間：99 年 12 月 26 日至 100 年 01 月 22 日止								
縣(市)代號			鄉(鎮市區)代號			普查區代號		
6	3		0	5		0	0	1
臺北市			中正區					
普查區內村里個數：01								
普查區內有人設籍總戶數(已扣除另查戶數)：92								
普查區內無人設籍地址總數：10								
普查員姓名：_____								

圖 2-3 普查名冊內頁

共 1 頁 ; 第 1 頁																				
戶號	戶內設籍人口數	戶籍戶長姓名	訪查結果			宅戶人數統計表				交表批次	訪查記錄					備註				
			書面填表	網路填表	未能訪查代號	宅戶	人口	新增	未能訪查		次數	1	2	3	4		5			
0021	001	3	吳真珍								1	月	日	/	/	/	/	/		
臺北市中正區南門里 20 鄰博愛路 202 巷 2 號											2	時間								
0022	001	2	李大年								1	月	日	/	/	/	/	/		
臺北市中正區南門里 12 鄰廣州街 4 號											2	時間								
0023	001	8	李小思								1	月	日	/	/	/	/	/		
臺北市中正區南門里 12 鄰廣州街 6 號											2	時間								
0024	001	2	郭小浩								1	月	日	/	/	/	/	/		
臺北市中正區南門里 12 鄰廣州街 8 號											2	時間								
0025	001	3	陳大智								1	月	日	/	/	/	/	/		
臺北市中正區南門里 10 鄰廣州街 8 號旁											2	時間								

未能訪查：1 屢訪未遇 2 全戶戶籍在本戶，但都沒有經常居住在本戶，3 此宅不存在、無人居住的其他 4 其他代號 / 拒訪 且本宅另有其他經常居住戶 房屋 (備註欄說明)

名冊封面除標明普查區代號外，另提供村里個數、設籍總戶數及無人設籍地址總數。名冊內頁則按門牌地址（宅）排序，若該地址具設籍人口，則逐一列出各設籍戶之人口數及戶長姓名等資訊；而若該地址無人設籍，則仍列於名冊內，惟僅列宅號。除門牌及戶籍資料外，每宅亦包含訪查結果、宅戶人數統計表、訪查記錄等欄位供普查員註記，以輔助實地訪查行政管理作業。

圖 2-4 普查名冊內頁宅、戶、人資料

戶籍基本資料欄		該地址在99年6月底以前之設籍情形	
宅號	戶號	戶內設籍人口數	戶籍戶長姓名
0001	001	4	王大明
臺北市中正區南門里12鄰廣州街2號			

如圖 2-4，為列印普查名冊宅、戶、人資料，必須由內政部戶役政資訊系統村里門牌資料及個人戶籍資料，針對門牌檔之每一門牌地址（宅），以門牌地址為鍵值連結戶籍檔以計算宅內各戶人口數，並逐一套印各宅每戶人口數及戶長姓名等資料，最後並總計普查區總戶數及未設籍地址數。

3. 普查區地圖：

為劃分個別普查員實地訪查責任範圍（普查區），平衡普查員工作負擔，並提供普查員責任範圍地圖資訊，主計處彙集各地籍圖資建置人口普查 GIS 系統，以孚前揭需求。人口普查普查區之劃分，以單一普查區包含 110 戶為原則，因此主計處先於 GIS 系統匯入各宅戶數資訊，再根據戶數原則及自然地理界線劃分普查區，並於實地訪查前列印地圖供普查員使用。普查區地圖樣式如圖 2-5。

圖 2-5 普查區地圖



4. 普查區街道範圍一覽表：

為利普查員實地訪查作業進行，主計處特產製普查區街道範圍一覽表，供普查員實地訪查時搭配普查區地圖參考。街道範圍一覽表提供普查區各路街或地區涵蓋之地址起號、訖號、設籍戶數及人數，並總計該普查區內設籍總戶數及人數於表頭。普查區街道範圍一覽表之樣式如圖 2-6。

圖 2-6 街道範圍一覽表

99年人口及住宅普查 普查區街道範圍一覽表											
高(三)級(縣市區)名稱			普查區代號			設籍總戶數		設籍總人數		普查區內村莊代號	
高(三)級(縣市區)代號			001			82		302		01	
高(三)級(縣市區)名稱			63			05				普查區內村莊名稱	
南門里											
路街名稱(含巷、弄、街、區、路)	起	弄	街	起	號	訖	號	單雙	設籍戶數	設籍人數	備註
廣井街				2	1	2		雙	18	60	
廣井街	4			1	1	1		單	2	9	
廣井街	4			8	1	8		雙	4	15	
廣井街	8			7	1	11		單	19	55	
廣井街	8	3		1	1	7	2	單	9	29	
廣井街	8	3		2	1	8		雙	5	15	
廣井街	8	5		1	1	7		單	2	9	
廣井街	8	5		2	1	10		雙	3	9	
博愛路	212			1	1	3	號之1	單	4	12	
博愛路	218			8	1	10		雙	13	49	
博愛路	224			1	1	3	號之2	單	6	19	
博愛路	224			2	1	6		雙	7	30	

為產製前述街道範圍一覽表，必須由門牌資料按普查區範圍區分，將普查區內各街道區段（區分至巷、弄、街等階層）分單、雙號標註該起號、訖號，並結合戶籍資料計算該街道區段所含設籍戶數及人數，最後並總計該普查區內設籍總戶數及人數。

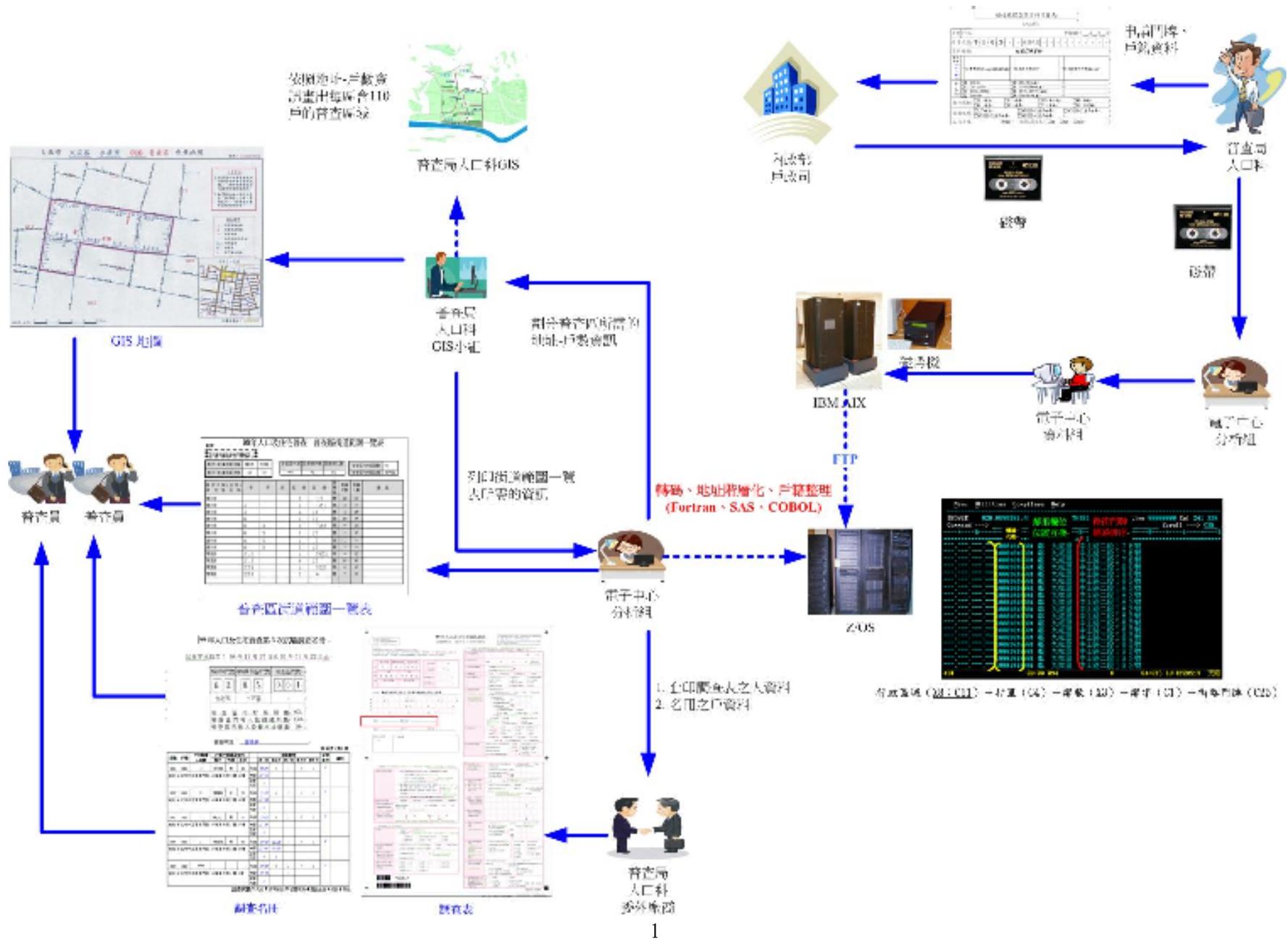
綜上所述，彙整人口普查實地訪查用表件用途及資料處理需求如表 2-1。

表 2-1 人口普查應用表件

表件	用途	資料處理需求
1. 套印普查表	於調查表套印各戶戶籍地址、戶內人口姓名、性別、出生年月日及身分證號後 5 碼，以輔助實地訪查作業	由戶籍資料針對每一地址區分不同設籍戶，再於該戶調查表逐一套印戶內人口各項資料
2. 普查區普查名冊	提供普查區宅戶資料，供普查員實地訪查參考，並提供訪查註記欄位以利訪查行政管理作業進行	針對門牌資料每一地址，以地址欄位為鍵值連結戶籍資料，計算每宅戶數及各戶人口數，並總計該普查區內之總戶數及未設籍地址數
3. 普查區地圖	標記個別普查員實地訪查責任範圍	利用戶役政資訊系統所提供之門牌及戶籍資料，搭配地籍圖資之天然地理界線，框定內含 110 戶之普查區範圍
4. 普查區街道範圍一覽表	提供普查區內各路街或地區所涵蓋之地址起號、訖號、設籍戶數及人數，供普查員實地訪查時參考	按普查區範圍就各街道區段分單、雙號標註該區段之起、訖號，並結合戶籍檔計算該區段所含設籍戶數及人數，總計該區設籍總戶數及人數
5. 外國人調查名冊	提供外國人口名冊及所屬普查區等資訊供調查員參考	利用內政部入出國及移民署外僑人口檔，以其地址欄位為鍵值連結本國人口母體資料，取得所屬普查區，並列印普查區外國人口清冊

以上作業皆仰賴普查母體之建置始能完成，而建置完成之母體資料則以循序文字檔（plain text sequential file）的格式儲存於大型主機平台，其整體建立過程如圖 2-7 所示。首先，由主計處普查局行文至內政部戶政司索取最新之戶役政資訊系統門牌及戶籍資料媒體檔，以磁帶儲存媒體的形式交由本中心利用磁帶機讀入大型主機，再經中文轉碼、地址階層化及戶籍資料整編等資料處理程序，透過各種交換格式提供委外廠商或其他系統列印或使用，最後這些經各項資料處理流程產出之表件會交予基層普查員，據以進行普查實地訪查作業。

圖 2-7 人口普查母體建置資料處理流程示意圖



第三節 人口普查母體資料來源及交換格式

如第二章第一節所述，人口普查本國人口母體資料來源為內政部戶役政資訊系統之村里門牌資料及個人戶籍資料，外國人口母體來源為內政部入出國移民署提供之外僑人口檔；惟人口普查母體建置及應用過程尚包含其他檔案交換，例如母體資料匯入 GIS 系統以劃分普查區及街道範圍一覽表、委外列印普查名冊及套印調查表。人口普查母體資料來源及交換格式彙整如表 2-2。

表 2-2 人口普查母體資料來源及交換格式一覽表

表件	資料產製	資料匯入	說明
1. 村里門牌資料	內政部	本中心	由內政部提供最新之村里門牌資料供人口普查母體建置
2. 個人戶籍資料	內政部	本中心	由內政部提供最新之個人戶籍資料供人口普查母體建置
3. GIS 普查區劃分匯入	本中心	主計處	將建置完成之人口普查母體資料匯入 GIS 系統，以按戶數搭配自然地理界線劃分普查區
4. GIS 普查區劃分匯出	主計處	本中心	將 GIS 劃分完成之普查區資訊匯出，回饋予人口普查母體
5. GIS 街道範圍一覽表	主計處	本中心	將 GIS 劃分完成之普查區街道範圍一覽表匯出，交由本中心進行大量列印
6. 普查名冊委外列印	本中心	委外廠商	提供委外廠商列印普查名冊，含普查區各宅每戶人口數及戶長姓名、總戶數及未設籍地址數等資料
7. 調查表委外套印	本中心	委外廠商	提供委外廠商於調查表上套印戶籍地址、戶內人口姓名、性別、出生年月日及身分證號後 5 碼等資料

而各來源資料及交換檔案格式如下：

1. 村里門牌資料（檔案長度：200 位元組）

欄位名	長度	起	迄	說明
保留區	60	1	60	
內政部縣市鄉鎮代號	8	61	68	

欄位名	長度	起	迄	說明
村里	16	69	84	E U C 中文
鄰編號	3	85	87	
鄰	4	88	91	E U C 中文
路街門牌	100	92	191	E U C 中文
編訂日期	7	192	198	
編訂類別	1	199	199	
NEWLINE	1	200	200	

2. 個人戶籍資料（檔案長度：819 位元組）

欄位名	長度	起	迄	說明
保留區	60	1	60	
身份證統一編號	10	61	70	
姓名	80	71	150	E U C 中文
出生年月日	7	151	157	
戶長身份證編號	10	158	167	
戶籍編號	8	168	175	
戶長姓名	80	176	255	E U C 中文
戶籍地址				
縣市鄉鎮代號	8	256	263	
村里名稱	16	264	279	E U C 中文
鄰別代號	3	280	282	
鄰	4	283	286	
路街門牌	100	287	386	E U C 中文
遷入年月日	7	387	393	
配偶姓名	80	394	473	E U C 中文
生父姓名	80	474	553	E U C 中文
生母姓名	80	554	633	E U C 中文
出生地代碼	5	634	638	
原住民身份代碼	1	639	639	
役別代碼	1	640	640	
養父姓名	80	641	720	E U C 中文
養母姓名	80	721	800	E U C 中文
與戶長關係	12	801	812	E U C 中文
婚姻狀況代碼	1	813	813	
教育程度代碼	2	814	815	

欄位名	長度	起	迄	說明
戶別代碼	1	816	816	
原住民識別代碼	1	817	817	
性別	1	818	818	
NEWLINE	1	819	819	

3. GIS 系統資料匯入格式 (檔案長度：306 位元組)

欄位名	長度	起	迄	說明
F01	1	1	1	
CITY_NAME	6	2	7	縣市名
DIST_NAME	8	8	15	鄉鎮市區名
F02	2	16	17	
LI_NAME	8	18	25	里名
ADDRESS	50	26	75	地址
F03	2	76	77	
ROAD	20	78	97	路名
SEC	12	98	109	段
SEC_STR	2	110	111	「段」字
LANE	20	112	131	巷號
LANE_STR	2	132	133	「巷」字
NUN	20	134	153	弄號
SUB_STR	2	154	155	「弄」字
LUN	8	156	163	街
NO	20	164	183	號
NO3	8	184	191	
NO_STR	2	192	193	「號」字
SPECIAL2	2	194	195	
NO4	6	196	201	
SPECIAL3	6	202	207	
NO5	6	208	213	
FLR_NO1	6	214	219	樓 (全形)
FLR_STR	2	220	221	「樓」字
FLR_DASH	2	222	223	「之」字
FLR_NO2	6	224	229	之號
ROOM	8	230	237	
ROOM_STR	2	238	239	

欄位名	長度	起	迄	說明
REMARK	14	240	253	
F08	1	254	254	
FLOOR	3	255	257	
F09	1	258	258	
DWELL_NEW	5	259	263	
F10	1	264	264	
ALO_ADDR	3	265	267	地址數
F11	1	268	268	
ALO_HOLD	5	269	273	戶數
F12	1	274	274	
ALO_PERSON	4	275	278	人數
F13	1	279	279	
DIST_CODE	4	280	283	鄉鎮市區代碼
LI_CODE	3	284	286	里代碼
LIN	3	287	289	鄰
F14	1	290	290	
SERIAL_NO	10	291	300	
F15	1	301	301	
DWELL_ORI	5	302	306	

4. GIS 系統資料匯出格式 (檔案長度：402 位元組，前 306 位元組同匯入格式)

欄位名	長度	起	迄	說明
F01	1	1	1	
CITY_NAME	6	2	7	縣市名
DIST_NAME	8	8	15	鄉鎮市區名
F02	2	16	17	
LI_NAME	8	18	25	里名
ADDRESS	50	26	75	地址
F03	2	76	77	
ROAD	20	78	97	路名
SEC	12	98	109	段
SEC_STR	2	110	111	「段」字
LANE	20	112	131	巷號
LANE_STR	2	132	133	「巷」字

欄位名	長度	起	迄	說明
NUN	20	134	153	弄號
SUB_STR	2	154	155	「弄」字
LUN	8	156	163	街
NO	20	164	183	號
NO3	8	184	191	
NO_STR	2	192	193	「號」字
SPECIAL2	2	194	195	
NO4	6	196	201	
SPECIAL3	6	202	207	
NO5	6	208	213	
FLR_NO1	6	214	219	樓(全形)
FLR_STR	2	220	221	「樓」字
FLR_DASH	2	222	223	「之」字
FLR_NO2	6	224	229	之號
ROOM	8	230	237	
ROOM_STR	2	238	239	
REMARK	14	240	253	
F08	1	254	254	
FLOOR	3	255	257	
F09	1	258	258	
DWELL_NEW	5	259	263	
F10	1	264	264	
ALO_ADDR	3	265	267	地址數
F11	1	268	268	
ALO_HOLD	5	269	273	戶數
F12	1	274	274	
ALO_PERSON	4	275	278	人數
F13	1	279	279	
DIST_CODE	4	280	283	鄉鎮市區代碼
LI_CODE	3	284	286	里代碼
LIN	3	287	289	鄰
F14	1	290	290	
SERIAL_NO	10	291	300	
F15	1	301	301	
DWELL_ORI	5	302	306	
ADDR_INDEX	60	307	366	

欄位名	長度	起	迄	說明
MATCH_CNT	1	367	367	對應次數
ED_CODE	4	368	371	
S_CODE	3	372	374	
ED_ID	4	375	378	普查區代碼
REGION	10	379	388	
JOIN_ED	10	389	398	
NODASH	4	399	402	號之"X"

5. 街道範圍一覽表資料檔 (檔案長度：226 位元組)

欄位名	長度	起	迄	說明
ED_ID	4	1	4	普查區代碼
CITY_NAME	6	5	10	縣市名
DIST_NAME	8	11	18	鄉鎮市區名
LI_NAME	8	19	26	里名
DIST_CODE	4	27	30	鄉鎮市區代碼
LI_CODE	3	31	33	里代碼
ROAD	20	34	53	路名
REGION	20	54	73	區域名
LANE	20	74	93	巷
NUN	20	94	113	弄
LUN	4	114	117	衛
START_NO	9	118	126	起號
END_NO	9	127	135	訖號
ODD_EVEN	1	136	136	單雙號辨識
ALO_PERSON	4	137	140	起訖間總人數
ALO_HOLD	4	141	144	起訖間總戶數
ALO_NO	4	145	148	起訖間總號數
ALO_ADDR	4	149	152	起訖間總地址數
ADDR_INDEX	50	153	202	
MARK	10	203	212	
ED_CODE	4	213	216	
JOIN_ED	10	217	226	

6. 普查名冊委外列印資料格式 (檔案長度：158 位元組)

欄位名	長度	起	迄	說明
縣市鄉鎮代碼	4	1	4	
FIL	1	5	5	
普查區代碼—列印	3	6	8	
FIL	1	9	9	
普查區註記				
普查區跨村里註記	1	10	10	
「(」	1	11	11	
普查區內村里個數	2	12	13	
「)」	1	14	14	
FIL	1	15	15	
宅號	4	16	19	
FIL	1	20	20	
戶號	3	21	23	
FIL	1	24	24	
戶內設籍人口數	4	25	28	
設籍戶數與無人設籍地址數之和	6	29	34	
普查區內有人設籍總戶數	6	35	40	
普查區內無人設籍地址總數	6	41	46	
FIL	1	47	47	
特殊對象註記				
外交部駐華使節辦公室	1	48	48	
安養中心	1	49	49	
外傭(看護工)	1	50	50	
姓名：EUC碼	24	51	74	
SO	1	75	75	
中文地址				
縣市名稱	6	76	81	
鄉鎮名稱	8	82	89	
村里名稱	8	90	97	
鄰號及鄰	8	98	105	
路街門牌號	50	106	155	
SI	1	156	156	
FLD-0D0A	2	157	158	

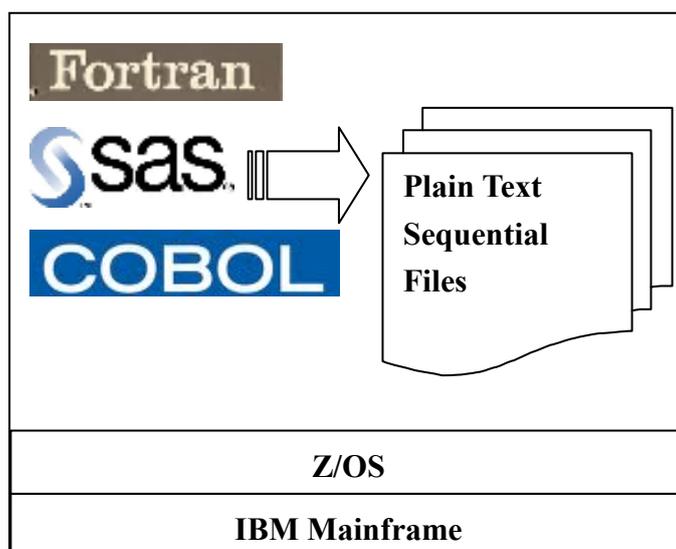
7. 調查表委外套印資料格式 (檔案長度：147 位元組)

欄位名	長度	起	迄	說明
縣市鄉鎮代碼	4	1	4	
FIL	1	5	5	
普查區代碼—列印	3	6	8	
FIL	1	9	9	
宅號	4	10	13	
FIL	1	14	14	
戶號	3	15	17	
FIL	1	18	18	
人口序號	4	19	22	(調查表該欄位固定印刷 1 2 3 4; 套印請依序號套印,第 5 人從第 2 張調查表 1 開始依序套印)
FIL	1	23	23	
戶內人口數	4	24	27	
FIL	1	28	28	
民前民國出生註記	1	29	29	
FIL	1	30	30	
出生年月日	6	31	36	
FIL	1	37	37	
性別註記	1	38	38	
FIL	1	39	39	
姓名：E U C 碼	24	40	63	E U C 碼
SO	1	64	64	
中文地址				
縣市名稱	6	65	70	
鄉鎮名稱	8	71	78	
村里名稱	8	79	86	
鄰號及鄰	8	87	94	
路街門牌號	50	95	144	
SI	1	145	145	
FLD-0D0A	2	146	147	

第四節 現行資料處理環境說明

現行人口普查母體建置資料處理作業由本中心支援，如圖 2-8 所示，硬體執行環境為 IBM 大型主機，作業系統為 Z/OS 390。人口普查母體資料以循序文字檔的形式儲存，本中心根據各項資料處理需求，於大型主機平台使用 Fortran、SAS 及 COBOL 等程式語言開發應用程式來進行母體建置相關資料處理作業。

圖 2-8 現行人口普查母體建置資料處理環境示意圖



第五節 現行中文轉碼及罕見字處理程序

因戶役政資訊系統中文編碼採 EUC 編碼，而現行人口普查母體資料處理平台為 IBM 大型主機平台之中文編碼採 NHC 編碼，故於母體建置作業進行前，需將來自戶役政資訊系統之地址中文欄位進行轉碼。目前所採行的方式為依據本中心歷年所建立的編碼對照表進行轉換，在大型主機平台上以 Fortran 語言開發之轉碼程式進行轉碼。目前轉碼對照表所包含的中文字近 1 萬 5 千餘字，若遇無對應編碼之罕見字則增補編碼對照表。

第六節 現行地址階層化及疑義地址處理程序

為增進普查員實地訪查之便利性，普查名冊依步行順序列印，即，依路街道、巷、弄、號順序分單雙號（先單後雙）列印。如圖 2-9 所示，以忠孝東路四段普查區為例：

1. 普查名冊先列印忠孝東路四段單號地址。列印忠孝東路四段 165 號、167 號... 至 179 號時遇到分隔忠孝東路四段 179 號及 183 號的忠孝東路 181 巷，在繼續訪查 183 號之前，普查員應先訪查 181 巷內之地址。

1.1 於是接續列印 181 巷內之地址，仍按先單後雙的原則先列印忠孝東路四段 181 巷內之單號地址。列印忠孝東路四段 181 巷 1 號、181 巷 3 號、181 巷 5 號後遇到分隔 181 巷 5 號及 181 巷 9 號的 181 巷 7 弄，在繼續訪查忠孝東路四段 181 巷 9 號之前，普查員應先訪查 181 巷 7 弄內之地址。

1.1.1 於是接續列印 181 巷 7 弄內之地址，仍按先單後雙的原則列印忠孝東路四段 181 巷 7 弄內的單號地址。

1.1.2 列印忠孝東路四段 181 巷 7 弄內的雙號地址。

1.2 列印忠孝東路四段 181 巷 9 號以後之單號地址。

1.3 列印 181 巷之雙號地址。列印忠孝東路 181 巷 2 號、181 巷 4 號...181 巷 8 號後又遇到分隔 181 巷 8 號及 181 巷 12 號的 181 巷 10 弄，在繼續訪查忠孝東路 181 巷 12 號前，普查員應先訪查 181 巷 10 弄內之地址。

1.3.1 於是先列印 181 巷 10 弄內之地址，仍按先單後雙的原則列印忠孝東路四段 181 巷 10 弄內的單號地址。

1.3.2 列印忠孝東路四段 181 巷 10 弄內的雙號地址。

1.4 接續列印忠孝東路四段 181 巷 12 號以後之雙號地址。

2. 接續列印忠孝東路四段 183 號以後之單號地址。

3. 列印忠孝東路四段之雙號地址。

圖 2-9 普查名冊列印順序示意圖



為孚上述需求，地址欄位須能區分「段」、「巷」、「弄」、「號」等階層，因此本中心訂定地址欄位階層標準格式並釋例如表 2-3。

表 2-3 現行地址階層標準格式

地址欄位階層	實例 1	實例 2	實例 3	實例 4	實例 5	實例 6
縣市名稱	台北縣	台北縣	台北縣	台北縣	台北縣	台北縣
鄉鎮名稱	鶯歌鎮	新店市	三芝鄉	新莊市	鶯歌鎮	新莊市
村里名稱	南靖里	頂城里	後厝村	瓊林里	中鶯里	雙鳳里
路（街、道）	文化路	安康路	土地公坑	瓊林路	和平街	中山路
段前引字串		—				三
【段】		段				段

巷前引字串	3 2 4	1 0 6		瓊英		
【巷】	巷	巷		巷		
弄前引字串	8			1 3		
【弄】	弄			弄		
衙前引字串	4					
【衙】	衙					
SP-CHAR1 : 特、臨、附		臨				
NO1 (全形數字)	5	2 3	5 6	2	2 9	5 3 8
DASH1 : 之、臨、附、一、~	之		之			
NO2 (全形數字)	1		1			
DASH2 : 之、臨、附、一、~			附			
NO3 (全形數字)			1 5			
【號】	號	號	號	號	號	號
SP-CHAR2 : 之、臨、附、頂		之		之		之
NO4 (全形數字)		1		3		1
SP-CHAR3 : 底、地下、地下之				附	地下	
NO5 (全形數字)				2		
主要樓層編號 (全形國字)			二		一	十一
【樓】			樓		樓	樓
【之】					之	
次要樓層編號 (全形數字)					9	
室前引字串						A
【室】						室

而戶役政資訊系統村里門牌資料之地址欄其各區段並無區分，如圖 2-10：

圖 2-10 戶役政資訊系統村里門牌資料地址欄位格式

807	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之14號	1501029020
808	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之15號三樓	1501029020
809	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之16號三樓	1501029020
810	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之17號四樓	1501029020
811	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之18號四樓	1501029020
812	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之19號四樓	1501029020
813	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之20號四樓	1501029020
814	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之21號四樓	1501029020
815	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之22號五樓	1501029020
816	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之23號五樓	1501029020
817	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之24號五樓	1501029020
818	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之25號五樓	1501029020
819	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之26號五樓	1501029020
820	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之27號地下一層	1501029020
821	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之28號	1501029020
822	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之29號	1501029020
823	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之30號	1501029020
824	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之31號	1501029020
825	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之32號	1501029020
826	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之33號	1501029020
827	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之34號	1501029020
828	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之35號	1501029020
829	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之36號	1501029020
830	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路325之37號	1501029020
831	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路327之10號五樓	1501029020
832	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路327之11號五樓	1501029020
833	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路327之12號五樓	1501029020
834	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路327之13號五樓	1501029020
835	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路327之14號三樓	1501029020
836	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路327之15號四樓	1501029020
837	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路327之16號四樓	1501029020
838	020	鄰	10015010	花蓮縣花蓮市	主力里	中華路327號	1501029020

若以圖 2-10「中華路325之27號地下一層」地址為例，為孚上述需求，便需將該地址切割為如表 2-4 所示。

表 2-4 現行地址欄位階層格式

地址欄位階層	實例
路（街、道）	中華路
段前引字串	
【段】	
巷前引字串	
【巷】	
弄前引字串	
【弄】	
街前引字串	
【街】	
SP-CHAR1：特、臨、附	
NO1（全形數字）	325
DASH1：之、臨、附、一、~	之
NO2（全形數字）	27
DASH2：之、臨、附、一、~	
NO3（全形數字）	
【號】	號

SP-CHAR2 :之、臨、附、頂	
NO4 (全形數字)	
SP-CHAR3 :底、地下、地下之	地下
NO5 (全形數字)	
主要樓層編號(全形國字)	一
【樓】	層
【之】	
次要樓層編號(全形數字)	
室前引字串	
【室】	

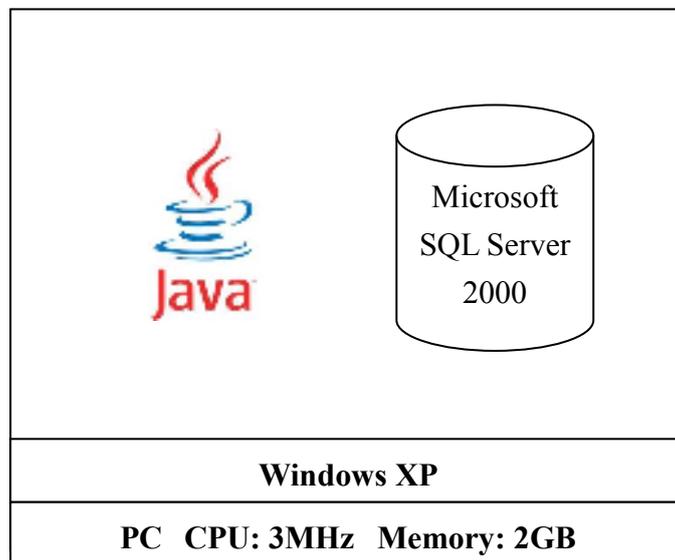
惟現行人口普查母體建置過程之地址階層化模組係於大型主機平台利用 Fortran 程式語言開發，因地址欄位包含許多錯別字或闕漏字，導致該模組需配合許多人工作業進行檢視排除，相當耗費人力。且各縣市地址欄位各具特性，格式繁多，導致該模組處理邏輯複雜，且須分縣市處理，造成必須依縣市別撰寫二十多支程式個別處理，使作業更形繁瑣。且大型主機環境欠缺友善的資料查詢介面，許多資料處理作業皆仰賴程式撰寫始能達成，形成本中心人力負荷。因此，本研究採用資料庫進行人口普查建置之可行性研究與實做，希望能簡化資料處理程序，增進查詢介面友善度，並減少人力耗費。

第三章 研究過程及成果

第一節 資料處理環境與範圍說明

如圖 3-1 所示，本研究採用 Microsoft SQL Server 2000 資料庫進行人口普查母體建置，硬體執行環境為個人電腦，作業系統為 Windows XP。整體環境為個人電腦安裝 SQL Server 資料庫軟體搭配 Java 程式語言進行資料處理。

圖 3-1 本研究資料處理環境示意圖



本研究之資料處理範圍，因時間限制，僅選定全國最大行政區域之臺北縣進行實作，即，採用資料庫針對臺北縣門牌及戶籍資料建置人口普查母體資料。

第二節 中文轉碼及罕見字處理流程

本研究採用資料庫進行人口普查母體建置的第一個步驟，是將戶役政資訊系統村里門牌及個人戶籍資料進行中文轉碼。目前較常見的中文編碼類型如表 3-1 所示，戶役政資訊系統係採用戶政 EUC-CSIC 中文編碼（以下稱 EUC 編碼），而因本研究之母體建置環境使用 Microsoft Windows 作業系統，故選用業界標準的 Big5 編碼進行儲存。因此，在將村里門牌資料及個人戶籍資料存入資料庫前，須先將戶政 EUC 編碼轉換為 Big5 編碼。

表 3-1 常見中文編碼

資訊系統	中文編碼	字集數	編碼方式	備註
MS Windows	BIG5/Unicode	2 萬以上	2~3 Bytes	
Linux	BIG5/Unicode	2.5 萬以上	1~4 Bytes	Utf-8 運作
戶政 EUC-CSIC	EUC-CSIC	7 萬以上	4 Bytes	
CNS 標準	CNS 11643	92,929 (96 年版)	3 Bytes	交換碼
業界標準	BIG5/BIG5E	13,053/3,954	2 Bytes	BIG5-2003

轉碼過程先將 EUC 編碼轉換為本中心「CNS11643 中文標準交換碼全字庫」（以下稱全字庫）網站的國家中文標準交換碼（CNS11643 編碼，以下稱 CNS 編碼），再利用全字庫所提供的 CNS 編碼與 Big5 編碼的對應轉換為 Big5 中文編碼。其中，EUC 編碼為全國戶役政單位使用，編碼長度為 4 個位元組，故能納編高達七、八萬的姓名用字。而 EUC 編碼的長度雖與 CNS 編碼不同，但卻採用 CNS 碼之編碼架構及所有字集，與 CNS 碼為一對一對應，其轉換規則如表 3-2 所示，EUC 之「8EA3A9FE」編碼會對應到 CNS 的「3-297E」編碼。其中，EUC 編碼的第 1 個位元組不轉換，第 2 個位元組則取 16 進位表示式的低位數作為 CNS 字面，亦即，16 進位的「A3」取「3」為 CNS 字面。第 3 個位元組則減 16 進位的「80」為 CNS 的高位元組，如 16 進位的「A9」減「80」後為「29」；第 4 個位元組同第 3 個位元組，減 16 進位的「80」後為 CNS 的低位

元組，如「FE」減「80」後為 16 進位的「7E」。以上便為 EUC 與 CNS 編碼之對應規則。

表 3-2 EUC 與 CNS 碼對應規則

EUC 位元組	EUC 與 CNS 碼對應規則	EUC	CNS
1	不轉換	8E	
2	CNS 字面 = EUC 第 2 個位元組 - A0h	A3	3
3	CNS 高位元組 = EUC 第 3 個位元組 - 80h	A9	29
4	CNS 低位元組 = EUC 第 4 個位元組 - 80h	FE	7E

轉為 CNS 編碼後，便可利用全字庫網站所維護的 CNS 與 Big5 對應表轉換為 Big5 編碼。其中會有無法對應至 Big5 編碼之罕見字，轉碼過程彙整這些罕見字供本中心進行 Big5 編碼造字。轉碼程式會記錄地址中無法轉換的 CNS 編碼，及其出現於村里門牌及個人戶籍資料地址欄位的次數，表 3-3 節錄出現較頻繁者。完整罕見字清冊彙整完成並由本中心進行 Big5 造字後，可將罕見字 CNS 與 Big5 對應表再以轉碼程式轉碼，完成罕見字處理。

表 3-3 地址常見罕見字

CNS編碼	出現次數	百分比	字形
03524A	27477	13.99%	館
033847	23242	11.83%	脚
0F4766	20548	10.46%	廊
033064	17732	9.03%	峯
03405A	16774	8.54%	塩
03215E	10743	5.47%	双
0F5067	10144	5.16%	礪
032739	5978	3.04%	邨
034C57	5728	2.92%	糠
033A28	4836	2.46%	厦
033C6B	3926	2.00%	温
034823	3835	1.95%	瑶
033E59	3704	1.89%	菓
032541	2976	1.52%	壘
0F3823	2035	1.04%	蔀

第三節 地址階層化處理流程

完成中文轉碼及罕見字處理後，便可將村里門牌及個人戶籍原始資料匯入資料庫，以進行地址欄位階層化處理。本研究所處理的臺北縣人口普查母體資料中，門牌地址總數為 1,680,413 筆，其中有人設籍的門牌地址數為 1,157,769 筆，約佔 67.9%，而無人設籍者則佔 30.7%。因人口普查需針對普查區內所有可能有人居住的住宅及處所進行調查，故原則上所有門牌地址皆必須進行訪查。

表 3-4 無人設籍門牌地址比率

資料來源	臺北縣	
有設籍地址數	1,157,769	67.9%
無人設籍地址數	522,644	30.7%
門牌地址總數	1,680,413	

為符合普查名冊按步行順序排序及分析普查區宅、戶、人數之需求，需將戶役政資訊系統村里門牌及個人戶籍資料地址欄進行階層化。現行階層化格式如第二章表 2-4 所示，但為簡化資料處理複雜度並提高地址階層清晰度，可針對現有階層格式進行簡併及調整如表 3-4，先簡併「號」相關階層但保留「NO1（全形數字）」階層，次簡併「樓」相關階層，然後獨立出「地區名稱」及「社區名稱」階層。調整後地址欄位階層格式如表 3-5。

表 3-5 簡併及調整地址欄位階層格式原則

調整	說明
合併「號」相關階層，但保留「NO1（全形數字）」階層	除因普查名冊及街道範圍一覽表的排序及起、訖號需求保留「NO1（全形數字）」階層外，餘合併為「號」階層
合併「樓」相關階層	
獨立「地區名稱」階層	如表 3-5 實例 5 臺北縣淡水鎮樹興里「興福寮 10 號」，原先「興福寮」被歸為「路（街、道）」階層，但明顯意義有區別，故新增「地

	區名稱」階層以提升階層清晰度
獨立「社區名稱」階層	如表 3-5 實例 6 臺北縣瑞芳鎮新峰里「三瓜子坑路國中新 50 號」，原先「國中新」被歸為「段前引字串」階層，但明顯意義有區別，故新增「社區名稱」階層以提升階層清晰度

表 3-6 簡併及調整後地址欄位階層格式

調整	階層名稱	實例 1	實例 2	實例 3	實例 4	實例 5	實例 6
	縣市名稱	台北縣	台北縣	台北縣	台北縣	臺北縣	臺北縣
	鄉鎮名稱	新店市	三芝鄉	鶯歌鎮	新莊市	淡水鎮	瑞芳鎮
	村里名稱	頂城里	後厝村	中鶯里	雙鳳里	樹興里	新峰里
新增階層	地區名稱					興福寮	
	路(街、道)	安康路	土地公坑	和平街	中山路		三瓜子坑路
	段前引字串	一			三		
	【段】	段			段		
	巷前引字串	106					
	【巷】	巷					
	弄前引字串						
	【弄】						
	街前引字串						
	【街】						
新增階層	社區名稱						國中新
合併「號」相關階層，保留「NO1(全形數字)」階層	SP-CHAR1：特、臨、附	臨					
	NO1(全形數字)	23	56	29	538	10	50
	DASH1：之、臨、附、一、~		之				
	NO2(全形數字)		1				
	DASH2：之、臨、附、一、~		附				
	NO3(全形數字)		15				
	【號】	號	號	號	號	號	號
	SP-CHAR2：之、臨、附、頂	之			之		
NO4(全形數字)	1			1			

合併「樓」相關階層	SP-CHAR3 : 底、地下、之			地下			
	NO5 (全形數字)						
	主要樓層編號 (全形國字)		二	一	十一		
	【樓】		樓	樓	樓		
	【之】			之			
	次要樓層編號 (全形數字)			9			
	室前引字串				A		
	【室】				室		
	註釋						

本研究除採用資料庫建置人口普查母體外，階層化資料處理則採用 Java 程式語言配合資料庫進行開發。因 Java 程式語言為物件導向語言，故細節之字串處理已封裝為 Java 程式語言套件 (package) 供直接使用。階層化資料處理本質上屬「正規表示法 (Regular Expression)」問題，Java 程式語言提供 `java.util.regex` 套件供開發使用，可讓開發者集中注意力於較高階的應用層次，提升問題解決效率，且經完整測試的套件品質高，維護成本較低。圖 3-2 為使用 `java.util.regex` 套件進行「號」階層切割的範例，將「SP-CHAR1 : 特、臨、附」、「NO1 (全形數字)」、「DASH1 : 之、臨、附、一、~」、「NO2 (全形數字)」、「DASH2 : 之、臨、附、一、~」、「NO3 (全形數字)」、「【號】」、「SP-CHAR2 : 之、臨、附、頂」、「NO4 (全形數字)」等階層以正規表示法表示出來，便可利用該套件將符合的字串取出。而戶役政資訊系統村里門牌地址經階層格式處理後之畫面如圖 3-3 所示。

圖 3-2 `java.util.regex` 「號」階層範例

```
// 先將「號」部份以正規表示法表示
java.util.regex.Pattern number_pattern =
    java.util.regex.Pattern.compile(
        "[特臨附0 1 2 3 4 5 6 7 8 9 甲乙A B之一]+號([之臨附][0 1 2 3 4 5 6 7 8 9]+)*");
// 於戶役政資訊系統門牌地址欄位尋找該正規表示式
java.util.regex.Matcher number_matcher = number_pattern.matcher(address);
// 若於地址欄尋得該正規表示式，便可利用套件功能逕行取出符合該正規表示式之字串
```

```

if (number_matcher.find()){
    number_part = address.substr(number_matcher.start(), number_matcher.end());
}

```

圖 3-3 戶役政資訊系統村里門牌地址階層化資料庫畫面

	縣市鄉鎮市區	村里	鄰	戶役政地址	地區名稱	路街道	段	巷	弄	街	社區名稱	號部分	樓部分
1	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷1之1號		三民路	二段	102巷				1之1號	
2	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷1之2號		三民路	二段	102巷				1之2號	
3	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷1之3號		三民路	二段	102巷				1之3號	
4	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷1之4號		三民路	二段	102巷				1之4號	
5	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷1號		三民路	二段	102巷				1號	
6	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷2之1號		三民路	二段	102巷				2之1號	
7	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷2之2號		三民路	二段	102巷				2之2號	
8	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷2之3號		三民路	二段	102巷				2之3號	
9	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷2之4號		三民路	二段	102巷				2之4號	
10	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷2號		三民路	二段	102巷				2號	
11	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷3之1號		三民路	二段	102巷				3之1號	
12	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷3之2號		三民路	二段	102巷				3之2號	
13	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷3之3號		三民路	二段	102巷				3之3號	
14	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷3之4號		三民路	二段	102巷				3之4號	
15	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷3號		三民路	二段	102巷				3號	
16	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷4之1號		三民路	二段	102巷				4之1號	
17	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷4之2號		三民路	二段	102巷				4之2號	
18	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷4之3號		三民路	二段	102巷				4之3號	
19	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷4之4號		三民路	二段	102巷				4之4號	
20	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷4號		三民路	二段	102巷				4號	
21	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷5之1號		三民路	二段	102巷				5之1號	
22	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷5之2號		三民路	二段	102巷				5之2號	
23	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷5之3號		三民路	二段	102巷				5之3號	
24	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷5之4號		三民路	二段	102巷				5之4號	
25	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷5號		三民路	二段	102巷				5號	
26	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷7之1號		三民路	二段	102巷				7之1號	
27	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷7之2號		三民路	二段	102巷				7之2號	
28	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷7之3號		三民路	二段	102巷				7之3號	
29	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷7之4號		三民路	二段	102巷				7之4號	
30	臺灣省臺北縣板橋市	九如里	006	三民路二段102巷7號		三民路	二段	102巷				7號	

第四節 疑義地址處理流程

因戶役政資訊系統村里門牌及個人戶籍資料包含許多錯別字地址，故須經各種方式過濾這些疑義地址，以提高母體資料品質，避免普查員據名冊及街道範圍一覽表上載之地址前往訪查時，卻發現無此地址的情形。本研究設計之疑義地址過濾方式如表 3-7。

表 3-7 疑義地址過濾方式

疑義地址過濾方式	實例	濾出原因說明
1. 關鍵字順序、格式或次數錯誤	<ul style="list-style-type: none"> ◆ 弘道街53巷9弄12號三樓 ◆ 五號路100號 ◆ 三界壇路109巷26之1三樓 	<ul style="list-style-type: none"> ◆ 「五號路」的「街」與「道」為同階層之關鍵字 ◆ 「五號路」中的「號」致程式誤判，但因出現兩個「號」而濾出 ◆ 正確地址須包含「號」關鍵字
2. 各階層加總不等於原地址	<ul style="list-style-type: none"> ◆ 仁化街菜寮商場10號 ◆ 中港路五守新村10號二樓 	<p>穿插於「路街道」、「巷」、「弄」、「街」「號」之間的社區名稱如「菜寮商場」或「五守新村」等無法由正規表示法處理，故以本規則進行過濾並建立對應表</p>
3. 各階層欄位過長或過短	<ul style="list-style-type: none"> ◆ 工二工業區工九路20號 ◆ 民權街鐵路巷10號 	<ul style="list-style-type: none"> ◆ 各階層欄位的長度過或過短者很可能為疑義地址 ◆ 「鐵路巷」中的「路」致程式誤判，但因「路(街、道)」階層欄位過長而濾出

1. 關鍵字順序、格式或次數錯誤：

有關「號」部分非正確格式及關鍵字出現兩次以上或順序錯誤的地址，經逐筆檢視更正後結果如表 3-8。

表 3-8 關鍵字順序、格式或次數錯誤地址人工處理

處理狀態	實例	處理說明	筆數
程式更正	<ul style="list-style-type: none"> ◆ 弘道街14號 ◆ 五號路232號 ◆ 市場二樓10號 	人工辨識特殊路名，設為參考檔後以程式整批處理	945
可更正	<ul style="list-style-type: none"> ◆ 蔦松二街87號15號 	逐筆檢視更正	254

	◆ <u>中正路西路</u> 691 號 ◆ 三界壇路 109 巷 <u>26 之 1</u> 三樓		
無法更正	◆ 雙菁路號 ◆ 新潭路三段	無法判斷「號」的值，故捨棄	41

2. 各階層加總不等於原地址：

有關各階層加總不等於原地址者，經檢視後發現主要為穿插於「路（街、道）」、「巷」、「弄」、「街」及「號」之間的社區名稱如「菜寮商場」或「五守新村」等，無法由正規表示法處理，故彙總對應表如表 3-10，並以程式根據對應表整批處理。

表 3-10 含「社區名稱」階層之地址

鄉鎮市區	原始地址實例	回組地址實例	筆數
臺灣省臺北縣新莊市	中港路 <u>五守新村</u> 10 號二樓	中港路 10 號二樓	332
臺灣省臺北縣瑞芳鎮	魚坑路 <u>瑞興新</u> 2 巷 15 號	魚坑路 2 巷 15 號	195
臺灣省臺北縣瑞芳鎮	中央路 48 巷 <u>吉祥園</u> 10 號	中央路 48 巷 10 號	133
臺灣省臺北縣瑞芳鎮	頂坪路 <u>瑞芳工業區</u> 110 號	頂坪路 110 號	68
臺灣省臺北縣瑞芳鎮	瑞濱路 <u>阿美家園</u> 40 號	瑞濱路 40 號	55
臺灣省臺北縣瑞芳鎮	明燈路一段 <u>瑞柑新村</u> 1 號	明燈路一段 1 號	48
臺灣省臺北縣瑞芳鎮	大埔路 <u>錢龍新城</u> 2 號	大埔路 2 號	45
臺灣省臺北縣瑞芳鎮	建基路一段 <u>弘祥新村</u> 20 號	建基路一段 20 號	42
臺灣省臺北縣瑞芳鎮	逢甲路 <u>裕民新城</u> 1 號	逢甲路 1 號	35
臺灣省臺北縣瑞芳鎮	大埔路 <u>吉慶公園城</u> 10 號	大埔路 10 號	265
臺灣省臺北縣瑞芳鎮	大埔路 <u>慶安新村</u> 40 號	大埔路 40 號	87
臺灣省臺北縣瑞芳鎮	大埔路 <u>台北新城</u> 10 號	大埔路 10 號	78
臺灣省臺北縣瑞芳鎮	大埔路 <u>有勝二村</u> 20 號	大埔路 20 號	64
臺灣省臺北縣瑞芳鎮	大埔路 <u>國宅新村</u> 40 號	大埔路 40 號	55
臺灣省臺北縣瑞芳鎮	大埔路 <u>東泰新村</u> 20 號	大埔路 20 號	55
臺灣省臺北縣瑞芳鎮	大埔路 <u>有勝新村</u> 10 號	大埔路 10 號	20
臺灣省臺北縣瑞芳鎮	三瓜子坑路 <u>民享新村</u> 10 號	三瓜子坑路 10 號	189
臺灣省臺北縣瑞芳鎮	三瓜子坑路 <u>正貿新村</u> 10 號	三瓜子坑路 10 號	119
臺灣省臺北縣瑞芳鎮	三瓜子坑路 <u>國中新</u> 20 號	三瓜子坑路 20 號	86
臺灣省臺北縣鶯歌鎮	重慶街 <u>南門商場</u> 10 號	重慶街 10 號	88
臺灣省臺北縣鶯歌鎮	中正三路 <u>鶯華新村</u> 42 號	中正三路 42 號	1
臺灣省臺北縣鶯歌鎮	中正三路 <u>二橋新村</u> 21 號	中正三路 21 號	7

臺灣省臺北縣鶯歌鎮	中正一路後村堰臨1號	中正一路臨1號	1
臺灣省臺北縣鶯歌鎮	中正三路318巷卓厝5號	中正三路318巷5號	5
臺灣省臺北縣鶯歌鎮	明圓街一棟1號... 明圓街九棟1號	明圓街1號	147
臺灣省臺北縣鶯歌鎮	欣欣街一棟1號... 欣欣街九棟1號	欣欣街1號	89
臺灣省臺北縣鶯歌鎮	晨曦街一棟1號... 晨曦街三棟1號	晨曦街1號	75
臺灣省臺北縣金山鄉	中山路新市場10號	中山路10號	24
臺灣省臺北縣金山鄉	金包里街舊市場1號	金包里街1號	8
臺灣省臺北縣三重市	仁化街菜寮商場10號	仁化街10號	10
臺灣省臺北縣三重市	重新路四段農場內2號	重新路四段2號	5
臺灣省臺北縣泰山鄉	仁愛路6巷泰安新村27號	仁愛路6巷27號	2
臺灣省臺北縣板橋市	四川路二段橋頭1之1號	四川路二段1之1號	2

另若「路」、「街」、「道」等關鍵字出現於「地區名稱」，亦可能誤判為「路（街、道）」階層，導致各階層加總不等於原地址。亦彙總對應表如表 3-11，以程式根據對應表整批處理。

表 3-11 「地區名稱」階層含其他階層關鍵字之地址

鄉鎮市區	原始地址	回組地址	筆數
臺灣省臺北縣八里鄉	長道坑10號	長道10號	75
臺灣省臺北縣淡水鎮	車路腳1號	車路1號	62

3. 各階層欄位過長或過短：

若各階層欄位的長度較長者，很可能為疑義地址，或因關鍵字混淆致程式誤判欄位階層。因此，分別針對各階層檢視字串長度，若過長或過短便以資料庫查詢語言查出檢視，以「地區名稱」階層為例，分析該階層字串長度分布如表 3-12。

表 3-12 「地區名稱」字串長度分布

「地區名稱」字串長	筆數	百分筆
7	37	0.00%

6	33	0.00%
5	49	0.00%
4	5846	0.34%
3	16970	1.00%
2	17596	1.03%
0	1663804	97.62%
總計	1704335	

其中，由於階層欄位字串長度 5 以上者顯著較少，故可能為階層化有誤地址。使用資料庫查詢語言查詢長度為 5 以上之地區名稱如表 3-13，經對照母體既存之地區名稱發現，林口鄉「粉寮醒吾新村」之「粉寮」應為地區名稱，而「醒吾新村」應為社區名稱；同上，雙溪鄉「外平林康福家園」之「外平林」為地區名稱，而「康福家園」為社區名稱；而其餘地區名稱為實際存在之地區名稱。此階段階層化有誤之地址存入社區名稱對照表以程式處理。

表 3-13 「地區名稱」字串長度大於 5 之地址

縣市鄉鎮市區	地區名稱	筆數
臺灣省臺北縣林口鄉	工二工業區	1
臺灣省臺北縣泰山鄉	坡雅頭信華	1
臺灣省臺北縣鶯歌鎮	宏德司法新村	5
臺灣省臺北縣林口鄉	粉寮醒吾新村	28
臺灣省臺北縣雙溪鄉	外平林康福家園	37
臺灣省臺北縣石門鄉	白沙灣別墅	47

次分析「路街道」階層字串長度分布，結果如表 3-14。

表 3-14 「路街道」字串長度分布

「路街道」字串長	筆數	百分筆
15	1	0.00%
8	58	0.00%
6	414	0.02%
5	3922	0.24%

4	203416	12.23%
3	1455230	87.48%
2	302	0.02%
1	84	0.01%

1663427

其中，由於階層字串長度 6 以上者或 2 以下者顯著較少，故可能為階層化有誤地址。使用資料庫查詢語言查詢字串長度 6 以上或 2 以下之路街道稱如表 3-15，經對照母體資料既存路街道名稱發現「工二工業區」應為地區名稱，而路街道名稱應為「工九路」；「民權街鐵路巷」的路街道階層應為「民權街」，「鐵路巷」則應為巷欄位；「街頂巷」應為巷欄位；且「弘道」應為地區名稱而非路街道欄位。此階段階層化有誤之地址經人工檢視後存入相關對照表以程式處理。

表 3-15 「路街道」字串長度大於 6 或小於 2 之地址

縣市	村里	地址	路街道
臺灣省臺北縣林口鄉	東林村	工二工業區工九路20號	工二工業區工九路
臺灣省臺北縣林口鄉	東林村	工二工業區工六路100號	工二工業區工六路
臺灣省臺北縣瑞芳鎮	龍興里	民權街鐵路巷10號	民權街鐵路
臺灣省臺北縣瑞芳鎮	龍興里	民權街鐵路巷11號	民權街鐵路
臺灣省臺北縣瑞芳鎮	崇文里	街頂巷100號	街
臺灣省臺北縣瑞芳鎮	崇文里	街頂巷101號	街
臺灣省臺北縣三峽鎮	弘道里	弘道89之5號	弘道
臺灣省臺北縣三峽鎮	弘道里	弘道89號	弘道

至於其他階層如「段」、「巷」、「弄」、「街」等，經統計及檢視分析後，皆無階層化有誤之情形。

第五節 人口普查母體資料庫品質

為了解本研究建置之人口普查母體資料庫品質，將本研究建置之母體資料與現行大型主機人口普查母體資料進行完整性及階層化比對。因現行大型主機人口普查母體階層格式與本研究不盡相同，故將階層格式按表 3-6 方式調整後進行勾稽，結果如表 3-16。勾稽成功比率高達 99.92%，而未能勾稽成功之地址類型，亦多為人工調整或階層化規則不同所致。

表 3-16 本研究母體資料庫與現行人口普查母體資料勾稽結果

勾稽結果	筆數	百分比
成功	1,676,337	99.92%
疑義地址調整方式不同	486	0.03%
階層化結果不同	572	0.03%
簡寫或錯別字	356	0.02%

其中疑義地址調整方式不同如「工二區工六路 100 號」之地址，本研究參酌臺北縣母體既有地區名稱將「工二區」修正為「工二工業區」，現行母體資料則未調整；又如「中山路 248 號地下一層 1 號」因出現兩個「號」關鍵字，故本研究調整為「中山路 248 號地下一層 1 室」，而現行母體資料亦未調整。故有極小比率之人工調整差異。另階層化結果不同者節錄如表 3-17。

表 3-17 本研究與現行母體階層規則不同類型

類型及實例	現行母體階層	本研究母體階層	筆數
無「路街道」但有「巷」： 烏勢巷 150 號	「烏勢巷」歸為「路街道」階層	「烏勢巷」歸為「巷」階層	4,77
「號」後接英文字母： 萬里加投 248 之 8 號 A	「A」歸為「樓」階層	「A」歸為「號」階層	34
只有「號」： 430 號	於「路街道」階層補上村里名稱	不做處理	19

此外，勾稽過程亦由現行母體資料發現戶役政資訊系統存在許多簡寫或錯別字的地址如表 3-18，這部份地址可匯總為對應表以程式處理。

表 3-18 簡寫及錯別字對照表

鄉鎮市區	範例地址	現行母體資料	筆數
臺灣省臺北縣鶯歌鎮	鶯桃路 9 1 巷 3 1 弄 1 街 6 號	鶯桃路 9 1 巷 3 1 弄 1 街 6 號	155
臺灣省臺北縣三芝鄉	蕃社后 2 0 號	番社后 2 0 號	108
臺灣省臺北縣金山鄉	泉路 1 0 號	溫泉路 1 0 號	70
臺灣省臺北縣金山鄉	秀峰坪 1 0 號	秀坪 1 0 號	10
臺灣省臺北縣金山鄉	牛埔子 1 0 號	牛埔仔 1 0 號	10
臺灣省臺北縣平溪鄉	紫橋坑 1 號	柴橋坑 1 號	3

綜上所述，與現行人口普查母體資料勾稽後發現，採用資料庫進行建置之母體資料品質符合需求。本研究經階層化處理之人口普查母體資料庫畫面如圖 3-4 所示，可用以產製普查名冊。

圖 3-4 本研究母體資料名冊資料庫畫面

--名、戶、人資料
 select A.townname as 縣市鄉鎮市區, A.villagename as 村里, A.neighbor_number as 鄰, 地區名稱, 路街道, 巷, 弄, 街, 社區名稱, 號部分 as 號, 樓部分 as 樓, A.address as 宅識別, B.HU_ID as 戶號, B.HU_INNER as 戶內人口, B.HU_HEAD_ID as 戶長姓名
 from H2O_UNION_ADDR\$01_D990623 A, H2O_HU99\$01_L819_D990623 B
 where A.status like 'OK%' and A.townname='臺灣省臺北縣中和市' and A.villagename='中原里'
 and A.towncode=B.towncode and A.villagename=B.villagename and A.address=B.address
 group by A.townname, A.villagename, A.neighbor_number, 地區名稱, 路街道, 巷, 弄, 街, 社區名稱, 號部分, A.address, 樓部分, B.HU_ID, B.HU_HEAD_ID
 order by A.townname, A.villagename, A.address, B.HU_ID

縣市鄉鎮市區	村里	鄰	地區名稱	路街道	巷	弄	街	社區名稱	號	樓	宅識別	戶號	戶內人口	戶長姓名
1	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				1 3 號		中山路二段 3 4 0 巷 1 3 號	F9049639	5	***
2	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				1 4 號		中山路二段 3 4 0 巷 1 4 號	FC673264	5	***
3	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				1 5 號		中山路二段 3 4 0 巷 1 5 號	FD126445	3	***
4	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				1 6 號		中山路二段 3 4 0 巷 1 6 號	FC674147	4	***
5	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				1 7 號		中山路二段 3 4 0 巷 1 7 號	FC673175	6	***
6	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				1 8 號		中山路二段 3 4 0 巷 1 8 號	FC446888	5	***
7	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				1 9 號		中山路二段 3 4 0 巷 1 9 號	FC673213	6	***
8	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				2 0 號		中山路二段 3 4 0 巷 2 0 號	F1659489	1	***
9	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				2 1 號		中山路二段 3 4 0 巷 2 1 號	F8371103	5	***
10	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				2 3 號		中山路二段 3 4 0 巷 2 3 號	F1156275	1	***
11	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				2 4 號		中山路二段 3 4 0 巷 2 4 號	FQ978062	1	***
12	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 0 號		中山路二段 3 4 0 巷 3 0 號	F0338685	5	***
13	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 1 號		中山路二段 3 4 0 巷 3 1 號	D1918992	3	***
14	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 4 號		中山路二段 3 4 0 巷 3 4 號	A9660957	4	***
15	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 5 號		中山路二段 3 4 0 巷 3 5 號	FS499625	1	***
16	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 5 號		中山路二段 3 4 0 巷 3 5 號	FS503916	1	***
17	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 5 號		中山路二段 3 4 0 巷 3 5 號	FS504432	1	***
18	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 5 號		中山路二段 3 4 0 巷 3 5 號	FS504459	1	***
19	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 7 號		中山路二段 3 4 0 巷 3 7 號	FP774056	2	***
20	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 8 號		中山路二段 3 4 0 巷 3 8 號	A3272122	6	***
21	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				3 9 號		中山路二段 3 4 0 巷 3 9 號	FQ980229	1	***
22	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				4 1 號		中山路二段 3 4 0 巷 4 1 號	Q2108388	5	***
23	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				4 2 號		中山路二段 3 4 0 巷 4 2 號	FA184398	5	***
24	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				4 4 號		中山路二段 3 4 0 巷 4 4 號	FM504574	2	***
25	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				4 4 號		中山路二段 3 4 0 巷 4 4 號	FR630361	1	***
26	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				4 5 號		中山路二段 3 4 0 巷 4 5 號	FC278314	1	***
27	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				4 7 號		中山路二段 3 4 0 巷 4 7 號	MO688617	2	***
28	臺灣省臺北縣中和市	中原里	011	中山路	3 4 0 巷				4 8 號		中山路二段 3 4 0 巷 4 8 號	FJ164183	5	***

方格 訊息
 查詢此次完成。 172.16.7.126 (8.0) sa (52) Test 0:00:22 1878 筆資料列 第 2 行, 第 1 欄

第六節 執行時間及人工參與分析

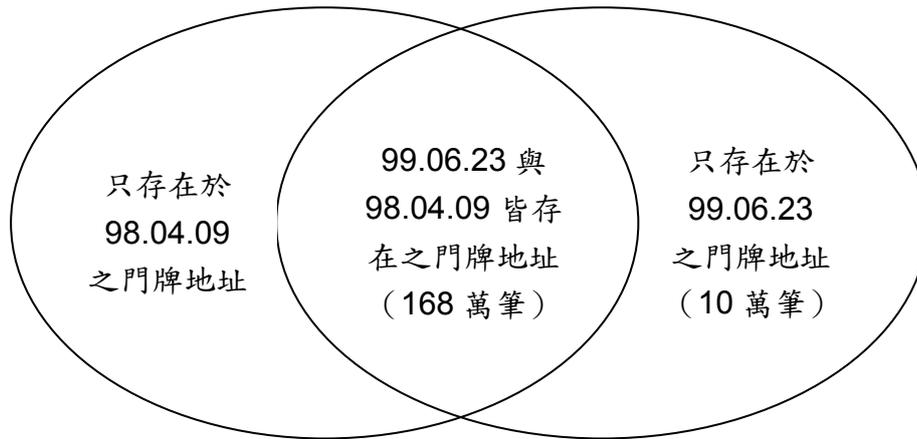
本研究採用資料庫進行台北縣人口普查母體建置之各階段執行時間及人工處理時間如表 3-19。

表 3-19 執行時間與人工參與分析表

處理階段	資料來源	自動化處理時間	人工參與	人工處理時間
轉碼與罕見字處理	村里門牌資料 (168 萬筆)	8 分鐘	整理 369 個罕見字 CNS 與本中心 Big5 造字編碼對應表	約 6 小時
	個人戶籍資料 (388 萬筆)	50 分鐘		
匯入資料庫	村里門牌資料 (168 萬筆)	15 分鐘	無	
	個人戶籍資料 (388 萬筆)	120 分鐘		
地址階層化	人口普查母體地址 (168 萬筆)	80 分鐘	無	
疑義地址處理	無		針對「關鍵字順序、格式或次數錯誤」、「各階層加總不等於原地址」、「各階層欄位長度過長或過短」等疑義狀況進行人工檢視排除	約 3 天

由表 3-19 可見主要時間還是在於人工處理時間，尤其是為提高母體資料品質所執行的疑義地址處理。但因本研究在設計資料庫綱要 (schema) 時，便已考慮此部份所需人工繁瑣耗時，且每次取得戶役政資訊系統最新村里門牌資料及個人戶籍資料時須能利用歷史成果簡化作業，因此在資料庫綱要設計時便設計原始地址保留欄位，搭配調整後地址欄位，每次取得最新村里門牌及個人戶籍資料時，可藉由比對該原始地址是否已存在於母體資料庫略去該部分地址，僅需針對異動部份做處理，大幅降低作業成本。

圖 3-5 門牌地址異動情形



以臺北縣為例，如圖 3-5 所示，99 年 6 月 23 日的村里門牌資料與 98 年 4 月 9 日相比，168 萬筆門牌地址中僅約 10 萬筆為新增或異動地址，故以 98 年 4 月 9 日版的村里門牌資料完成母體建置後，對於 99 年 6 月 23 日版之較新門牌地址僅須處理 10 萬筆新增地址，亦即 6% 的新增地址即可，大幅降低作業時間。

第七節 外國人口母體建置—地址連結應用

本章前揭部份皆在探討採用資料庫進行本國人口母體建置過程，惟人口普查範圍尚包括外國人口，此部份需求可利用建置完成之本國人口母體資料來確認外國人口資料來源（內政部入出國移民署外僑人口檔）地址正確性，並歸屬於劃定之普查區，據以進行調查行政作業相關規劃應用。

除人口普查之外國人口調查外，目前工商及服務業普查及農林漁牧業普查皆各自蒐集相關官方登記資料，結合歷次調查名單以建立正式普查作業母體，並據母體名冊地址規劃人力與作業以遂行調查。而工商及農林漁牧普查所蒐集之個別官方登記資料無法完整涵蓋母體整體範圍，故需結合多個資料來源比對出整體聯集作為完整母體範圍；惟因各登記資料之行政區域欄位可能有所闕漏，加以地址格式未盡一致，故於資料確認及比對上多借助人口普查母體地址資料以補足行政區域欄位，或藉以確認街道正確性。此時所衍生某資料來源地址與另一資料來源地址比對的需求，稱為地址連結；惟因各資料來源地址格式不一，且常有錯別字

及罕見字，故亟需較有效率的方法來進行連結。

以人口普查 GIS 系統為例，其地址資料來源與戶役政資訊系統不同，故 GIS 系統在接受戶役政資訊系統地址戶數資料時，便需進行地址連結。在此，本研究以外國人口地址連結為例，說明採用資料庫處理之實作過程。如圖 3-5，外國人口調查之母體資料來源為內政部入出國移民署之外僑人口檔，在調查進行前必須藉由其「居留地址」欄位判定外國人口所屬普查區，規劃及分配每位調查員責任範圍，再列印名冊交由調查員據以前往調查。

圖 3-5 內政部入出國及移民署外僑人口檔畫面

1	A	B	C	E	F	G	H	J	L	M	N	P
編號1	縣市鄉鎮	居留地址	性別	出生日期	居留證號	效期截止日期	英文姓名	國籍	職業	居留事由	服務處所/就讀學校/依親對象	
2	B00001	1119	臺南縣善化鎮*****	女	*****	*****	*****	211	18	永居(居住二十年者)	*****	
3	B00002	6304	臺北市中山區*****	男	*****	*****	*****	315	16	永居(連續居留七年)	*****	
4	B00003	6408	高雄市苓雅區*****	男	*****	*****	20130415	315	10	投資	*****	
5	B00004	0103	臺北縣中和市*****	男	*****	*****	20111203	210	34	投資	*****	
6	B00005	6303	臺北市大安區*****	女	*****	*****	20101106	315	10	投資	*****	
7	B00006	1904	臺中市西區*****	男	*****	*****	20110905	232	10	投資	*****	
8	B00007	6304	臺北市中山區*****	男	*****	*****	20110506	022	10	投資	*****	
9	B00008	0121	臺北縣三芝鄉*****	女	*****	*****	20121219	315	34	投資	*****	
10	B00009	0121	臺北縣三芝鄉*****	女	*****	*****	20121222	315	03	投資	*****	
11	B00010	6304	臺北市中山區*****	男	*****	*****	20120711	315	24	投資	*****	
12	B00011	2002	嘉義市西區*****	男	*****	*****	20110814	211	10	投資	*****	
13	B00012	6304	臺北市中山區*****	男	*****	*****	20120203	015	10	投資	*****	
14	B00013	6403	高雄市左營區*****	男	*****	*****	20120925	232	10	投資	*****	
15	B00014	6311	臺北市士林區*****	男	*****	*****	20130104	019	34	投資	*****	
16	B00015	6310	臺北市內湖區*****	男	*****	*****	20120711	019	10	投資	*****	
17	B00016	6303	臺北市大安區*****	男	*****	*****	20101216	009	34	投資	*****	
18	B00017	0301	桃園縣桃園市*****	男	*****	*****	20130326	019	34	投資	*****	
19	B00018	6301	臺北市松山區*****	男	*****	*****	20110314	212	05	投資	*****	
20	B00019	1906	臺中市內湖區*****	男	*****	*****	20111013	240	30	投資	*****	
21	B00020	6303	臺北市大安區*****	女	*****	*****	20110728	303	02	投資	*****	
22	B00021	6304	臺北市中山區*****	男	*****	*****	20110527	211	24	投資	*****	
23	B00022	6301	臺北市松山區*****	男	*****	*****	20110417	010	24	投資	*****	
24	B00023	6301	臺北市松山區*****	男	*****	*****	20120504	027	10	投資	*****	
25	B00024	0106	臺北縣新店市*****	男	*****	*****	20130211	027	34	投資	*****	
26	B00025	6311	臺北市士林區*****	男	*****	*****	20110118	027	10	投資	*****	
27	B00026	0310	桃園縣平鎮市*****	男	*****	*****	20101121	023	34	投資	*****	
28	B00027	6303	臺北市大安區*****	男	*****	*****	20110117	008	10	投資	*****	
29	B00028	6308	臺北市文山區*****	男	*****	*****	20110524	312	20	投資	*****	
30	B00029	6301	臺北市松山區*****	男	*****	*****	20110518	008	10	投資	*****	
31	B00030	1802	新竹市北區*****	男	*****	*****	20110917	015	10	投資	*****	
32	B00031	1907	臺中市南屯區*****	男	*****	*****	20110501	019	24	投資	*****	
33	B00032	6304	臺北市中山區*****	男	*****	*****	20110408	008	34	投資	*****	
34	B00033	0305	桃園縣蘆竹鄉*****	男	*****	*****	20120102	225	24	投資	*****	
35	B00034	1008	嘉義縣六腳鄉*****	男	*****	*****	20111107	014	10	投資	*****	

外僑人口檔的中文編碼為 Big5，故無須進行中文轉碼。將原始資料匯入資料庫後，便可與本國人口母體資料進行地址欄位完整比對。惟為提高地址連結成功率，須先針對外僑人口檔地址欄位進行格式整理。以「苗栗縣苗栗市建功里縣府路1之1號2樓」該地址為例，步驟如下：

1. 將半形數字轉為全形
2. 將縣市鄉鎮等行政區域名稱切出
 - ◆ 切割為「苗栗縣苗栗市」與「建功里縣府路1之1號2樓」
3. 將村里、鄰等行政區域名稱切出
 - ◆ 切割出「建功里」，剩下「縣府路1之1號2樓」
4. 依戶役政資訊系統規則，將「段」或「樓」字前的數字轉為國字
 - ◆ 將「縣府路1之1號2樓」調整為「縣府路1之1號二樓」

接著由程式進行連結時，將「1之1號」與「1號之1」兩種格式視為等同，以提高連結率。經上述處理後，86,334筆外國人口原始資料中共有80,976筆可成功連結至本國人口母體之地址資料（戶役政資訊系統村里門牌資料），連結成功率達93.8%，而連結成功者可據以判定所屬普查區。

剩下無法連結成功的5,358（6.2%）筆外國人居留地址，若以人工逐筆利用戶役政資訊系統網站進行判別，所需人力相當可觀。故本研究利用字串模糊比對演算法提供較接近之戶役政門牌地址，以利快速判別該居留地址所對應之戶役政門牌地址區段。如圖3-6，針對無法完整對應至戶役政門牌地址之居留地址，本研究經模糊比對後，列出本國人口母體地址資料中10筆相似度最高的地址區段，以利快速判定正確地址及所屬普查區。

圖 3-6 字串模糊比對推薦連結地址

原始地址	推薦地址	相似度
112-AC02276292-臺北縣中和市號自立路29巷14弄28之4號	臺北縣-中和市 -14149-台北縣中和市自立路29巷14弄[2~30]	0.600
	臺北縣-中和市 -14148-台北縣中和市自立路29巷[2~12]	0.486
	臺北縣-中和市 -14941-台北縣中和市連城路285巷14弄[2~8]	0.400
	臺北縣-中和市 -14975-台北縣中和市連城路319巷14弄[2~8]	0.400
	臺北縣-中和市 -16242-台北縣中和市永和路202巷14弄[2~2]	0.400
	臺北縣-中和市 -12767-台北縣中和市圓通路369巷14弄[2~24]	0.390
	臺北縣-中和市 -14283-台北縣中和市景平路278巷14弄[2~14]	0.390
	臺北縣-中和市 -16770-台北縣中和市保健路29巷[2~2]	0.389
	臺北縣-中和市 -14239-台北縣中和市景平路278巷14弄[16~16]	0.381
	臺北縣-中和市 -16318-台北縣中和市連城路285巷14弄[10~12]	0.381
	114-FD00450653-臺北縣中和市民德街146巷5弄3號5樓	臺北縣-中和市 -12055-台北縣中和市民德路146巷5弄[1~7]
臺北縣-中和市 -12054-台北縣中和市民德路146巷[1~3]		0.514
臺北縣-中和市 -12600-台北縣中和市民利街14巷5弄[1~15]		0.486
臺北縣-中和市 -12057-台北縣中和市民德路146巷7弄[1~13]		0.474
臺北縣-中和市 -12173-台北縣中和市民德路25巷5弄[1~5]		0.444
臺北縣-中和市 -15990-台北縣中和市新生街18巷5弄[1~3]		0.444
臺北縣-中和市 -12418-台北縣中和市民利街100巷5弄[1~3]		0.432
臺北縣-中和市 -15723-台北縣中和市立德街26巷5弄[1~17]		0.432
臺北縣-中和市 -16070-台北縣中和市新生街18巷5弄[1~39]		0.421
臺北縣-中和市 -16271-台北縣中和市民享街116巷25弄[1~1]		0.421
117-AD02330962-臺北縣永和市和永貞路94號5樓	臺北縣-永和市 -17051-台北縣永和市永貞路[2~8]	0.444
	臺北縣-永和市 -16984-台北縣永和市永貞路[54~68]	0.414
	臺北縣-永和市 -18708-台北縣永和市永貞路[72~84]	0.414
	臺北縣-永和市 -18710-台北縣永和市永貞路[12~52]	0.414
	臺北縣-永和市 -18696-台北縣永和市永貞路[90~142]	0.400
	臺北縣-永和市 -18977-台北縣永和市永貞路23巷[2~6]	0.400
	臺北縣-永和市 -16966-台北縣永和市永貞路[144~154]	0.387
	臺北縣-永和市 -16967-台北縣永和市永貞路156巷[2~2]	0.387
臺北縣-永和市 -16985-台北縣永和市永貞路70巷[2~24]	0.387	

以「臺北縣中和市號自立路29巷14弄28之4號」為例，「中和市」與「自立路」間之「號」字應為輸入錯誤，導致該地址無法連結成功，但由模糊比對推薦的地址區段可迅速判別此錯誤，判定該地址應屬於「14149-台北縣中和市自立路29巷14弄[2~30]」這個街道區段，而該區段之普查區編號為14149。

同上，則可藉由最相似於「臺北縣中和市民德街146巷5弄3號5樓」之10個地址區段發現，中和市應只有民德「路」而沒有民德「街」，故民德「街」應為筆誤，而該地址所屬普查區為12055。同理，亦可快速發現「臺北縣永和市和永貞路94號5樓」永貞路前之「和」字應為贅字，並可快速判定該地址所屬之普查區代號為18696。

以上便為本研究採用資料庫並利用地址連結完成外國人口母體建置之流程，除可減少人力耗費，本連結方式亦可供其他普查應用。

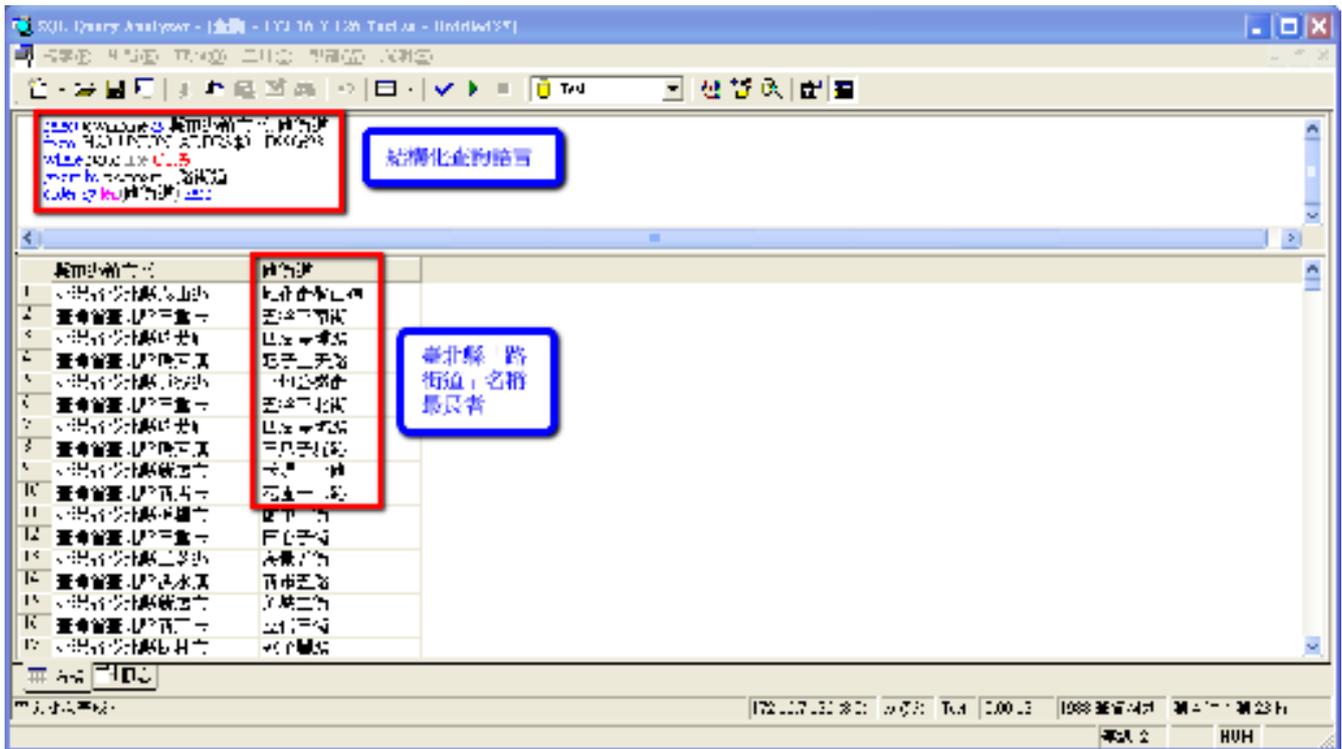
第四章 結論與建議

第一節 結論

1. 採用資料庫建置普查母體資料，可增進查詢親和性並減少程式開發負擔。

由於資料庫可使用結構化查詢語言標準進行查詢，且視窗環境較大型主機平台操作介面友善，因此可增進查詢親和性。另結構化查詢語言可支援彈性較廣之查詢或更新資料方式，以往在大型主機平台須開發應用程式始能達成之需求，有許多可使用結構化查詢語言便可達成，故可節省應用程式開發負擔。例如，若想了解臺北縣地址中「路（街、道）」階層名稱最長者，便可利用資料庫查詢語言逕行查詢如圖 4-1，而不須另行撰寫程式。

圖 4-1 以資料庫查詢語言查詢臺北縣路街道名稱最長者

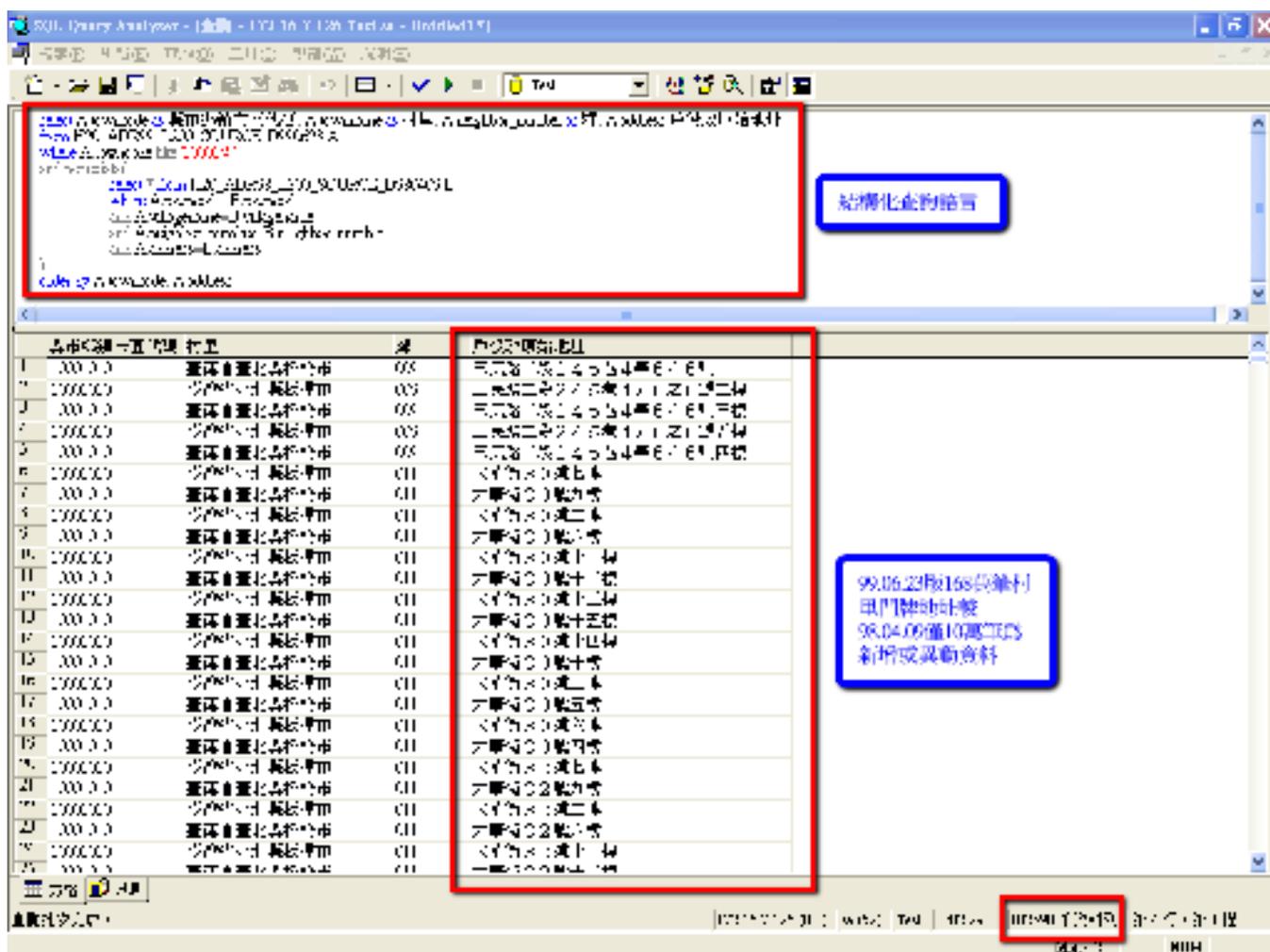


2. 本研究母體建置可簡化公務檔案處理流程、提升效率並減少人力耗費。

本研究採用資料庫建置普查母體資料，配合簡併及調整地址欄位階層格式，

及使用 Java 程式語言 java.util.regex 正規表示式套件進行階層化，可減低公務檔案地址處理複雜度，減少應用程式數量及開發負擔。另因本研究資料庫綱要同時保留原始地址欄位及調整後地址欄位，故母體資料建置完成後，便可以其為基底，僅針對異動部份處理。如圖 4-2，臺北縣 99 年 6 月 23 日之 168 萬筆村里門牌資料中，僅 10 萬筆為較 98 年 4 月 9 日後之新增或異動地址，故僅需處理此 6% 的新增地址，大幅提升效率並減少人力耗費。

圖 4-2 新增或異動地址比對



3. 採用資料庫進行疑義地址處理，可兼顧母體資料品質與效率。

本研究第三章第四節使用資料庫進行疑義地址處理，因可使用資料庫查詢語言快速比對「各階層加總不等於原始地址」或「各階層欄位過長或過短」等疑義地址，並逕行欄位更正，故可同時兼顧母體資料品質與效率。且經與現行母體資

料進行勾稽比對後，如表 3-16 所示，勾稽成功率高達 99.92%，母體資料庫品質符合需求。

4. 可以人口普查母體資料庫為基底，建置模糊比對功能供其他調查連結地址。

如第四章第七節所述，各普查母體建置過程因各登記資料來源之行政區域欄位可能有所闕漏，加以地址格式不一致且常有錯別字及罕見字，故亟需較有效率的方法來與戶役政村里門牌資料進行比對。此時可使用本研究搭配資料庫索引及字串模糊演算法之方式，快速推薦村里門牌資料最近似的門牌地址，大幅降低為提高連結率所耗費之時間及人力成本。而較高之連結率則可使調查母體資料品質提高，並使連結公務檔案所得普查資訊有更高之正確性。

第二節 建議

1. 建立母體資料庫平台及雛型系統。

由於本研究結果顯示採用資料庫進行母體建置具有親和及效率上的優點，故建議擇定實際普查作業建立雛型系統，導入資料庫處理作業模式，進而透過教育訓練，推廣至三大普查母體建置。而因工商母體資料為許多調查之母體來源，資料更新週期較短，故建議以工商母體為雛型系統標的，逐步完善工商母體資料庫建置及相關規畫。

2. 建立母體資料庫資通安全機制。

因普查母體資料包含敏感性資料如姓名、身分證字號、地址、收入等，故若欲使用資料庫進行母體資料建置，須妥善規劃資通安全機制，包含權限管理、存取稽核紀錄及營運持續性計畫等，始能保障個人資料與組織業務運作安全。

3. 母體資料庫建立後，可整體規劃普查商業智慧環境。

普查母體資料庫建置完成後，除名冊列印等前置作業外，實際訪查作業後續之調查資料處理、公務檔案連結、調查資料檢誤、分析以及結果表編製與列印皆屬整體作業可搭配商業智慧解決方案進行整合規劃，可導入萃取轉製載入(ETL)工具支援公務檔案處理，並結合既有之光學字元識別(OCR)自動化調查資料處理，且整合資料倉儲(Data Warehousing)、線上分析處理(OLAP)及報表編製軟體使普查價值進一步提升。

參考文獻

- [1] 行政院主計處「99年人口及住宅普查-普查作業手冊」。
- [2] 行政院主計處電子處理資料中心全字庫網站，<http://www.cns11643.gov.tw>。
- [3] 行政院內政部戶政司全球資訊網，<http://www.ris.gov.tw/>。
- [4] 臺北市政府工務局臺北市地理資訊 e 點通，<http://addr.tcg.gov.tw/>。