

資訊技術在戶口及住宅普查上之應用與發展

資訊技術在普查運用上，若能從資料源頭，建立可完整結合的共享與共通機制，將使各種檔案資料在資訊技術輔助下之運用更有效率、更具彈性，且大幅提升其運用範疇與效益。

◎羅國華、劉惠玲、周元暉（行政院主計處第四局研究委員、專員、研究員）

壹、前言

戶口及住宅普查是基本國勢調查之一，世界各國都相當重視，惟其作業繁雜且資料量龐大，須用大量經費與人力，因此如何有效運用適當方法技術，以提升普查整體效益相當重要。隨著資訊技術的發展，使得在個人電腦環境的普及精進、區域網路(LAN)及網際網路(Internet)的盛行後，對於普查作業已造成重大變革。有鑑於此，本次普查規劃乃積極進行多項資訊技術興革，主要包括光學字元辨識系統(OCR)處理環境的建置、連結多種公務登記檔案產生普查資訊及檢核普查相關資料、運用地理資訊(GIS)技術在直轄市建構數值圖檔運用、普查行政作業電子系統的建置、利用網際網路進行人員溝通及資料訊息的傳遞、以及主機系統運用方法的突破等，以增進普查作業效率及資料應用範疇。各項資訊作業，除OCR系統委由工業技術研究院及民間業者開發外，其餘相關的作業規劃、方法設計、程式開發及資料連結運用等技術，均由本處人員自行建置完成。

貳、普查資訊技術發展概況

一、普查資訊技術主要發展趨勢

世界第一部數位式商用電腦 UNIVAC I 於 1951 年問世，即在美國普查局運作，五十餘年來隨軟硬體技術的精進，對普查作業不斷地產生變革影響。2000 年前後許多國家辦理普查，面對的是社會資訊化程度日深，調查環境日益惡化，不但經費減少，民眾配合意願亦逐漸降低，為改善困境，因此廣泛運用各種資訊技術在普查作業上。諸如北歐國家及荷蘭採行登記式(Register-Based)普查，運用資料倉儲(Data Warehouse)觀念與技術，大量連結公務登記資料，進行資料蒐集與整合作業，成效良好。亦有超過二十餘國開發 OCR 系統，或進而發展自動註號(Automatic Coding)及資料檢誤(Editing)能力，大幅提升普查資料處理效能。雖然地理資訊系統的發展常受限於地理圖檔的完備情形，但在聯合國的推展下已有五十國以上進行程度不等的運用，增進普查規劃作業品質，擴展普查資料應用範疇。此外新加坡更首創網際網路、電腦輔助電話訪問(CATI)、面訪三階段的普查方式，並連結整合公務登記資料，有效減少訪查費用，提升資料品質，因而電子式(E-Based)問卷正對傳統紙式(Paper-Based)問卷的調查方式形成衝擊。至於其他有關普查作業管理、品質保證檢驗、資料製表作業、普查資料

供應等方面，各國在小型化（Downsizing）潮流趨勢下，大幅減少主機（Mainframe）環境的作業方式，以網路為基礎發展電腦主從（Client-Server）架構系統輔助各項作業，資訊技術正快速的蓬勃發展，對傳統普查的觀念及方法產生衝擊，如何善用有限資源應是吾人所當慎思與努力者。

二、我國普查資訊技術應用情形

國內普查作業電腦化始於 1966 年普查時運用 IBM S360-30 作為資料輸入與處理；其後 1980、1990 年兩次普查仍延續運用主機環境，惟在作業上均進行許多改善，以縮短處理時程及降低費用，而能增進作業整體效率；1990 年普查更首次引進光學符號閱讀機（OMR），使傳統人工輸入資料方式產生變革，有效提升資料品質。

近十餘年來，隨科技的長足進展，國內資訊環境已在快速變遷，使得許多觀念與方法在運用上得以擺脫環境限制而能逐步推展。檢視 2000 年普查，更勝以往地運用各項資訊技術於普查各階段作業上（表一），除下節介紹的三項主要興革外，包括個人電腦環境的廣泛運用，結合網際網路進行訊息的溝通及傳遞，利用 Visual Basic 及 Access 開發「行政電子作業系統」促進行政作業的標準一致化，並為搭配主機檢誤以 Access 開發「批次檢誤更正系統」；而在傳統的主機作業中，引進「磁帶館儲存管理系統」、「OnDemand 報表管理系統」等，並改善程式發展及系統資源分配環境，以因應處理普查大量資料之所需。

表一 2000 年戶口及住宅普查各階段使用資訊技術情形

	工作項目	使用軟體種類	使用硬體種類
規劃設計階段	1.蒐集參考資料	1.瀏覽器 2.電子郵件	1.個人電腦 2.網路通訊設備
	2.編製普查文件	1.WORD、EXCEL、漢書 2.POWER POINT、VISIO	
	3.開發行政作業電子系統	1.Visual Basic、Access 2.電子郵件	
調查前準備階段	1.編製普查底冊	1.SAS、COBLE、FORTRAN 2.FILE-AID/ABEND-AID/STROBE	1.IBM RS/6000 2.IBM S390 主機 3.IBM Info 3900 及 4000 高速印表機 4.個人電腦 5.彩色印表機 6.DAT 磁帶機
	2.套印普查表	3.AFP PRINTING 4.OnDemand 報表管理系統	
	3.劃分普查區	5.地理資訊系統	
實地調查階段		1.行政作業電子系統 2.電子郵件 3.瀏覽器	1.個人電腦 2.網路通訊設備
資料處理階段	1.資料輸入	1.光學字元閱讀辨識系統 2.大型資料庫（ORACLE） 3.防毒軟體 4.遠端監控軟體	1.IBM RS6000 伺服器 2.NT 伺服器/工作站 3.Kodak 9500/7500 高速掃描器及平台掃描器

	2.資料檢誤	5.資料庫備份軟體 6.主機檢誤系統 7.磁帶館儲存管理系統	4.DVD 光碟櫃 5.磁帶櫃、磁帶館 6.網路通訊設備 7.IBM S390 主機
	3.公務檔案取得及轉檔	1.WS-FTP 2.主機轉檔系統	1.IBM RS/6000 伺服器 2.IBM S390 主機
	4.資料整理及連結公務檔案	1.SAS、COBLE 2.FILE-AID/ABEND-AID/STROBE 3.OnDemand 報表管理系統 磁帶館儲存管理系統.	DAT 磁帶機 4.磁帶館 5.個人電腦 6.IBM Info 4000 高速印表機
	5.資料製表	1.SAS、COBLE 2.統計表製表系統 3.AFP PRINTING 4.WORD、EXCEL、漢書	
資料編 布階段	1.資料供應及查詢	1.SAS、COBLE 2.瀏覽器	1.個人電腦 2.CD 燒錄器
	2.資料保存	3.電子郵件 4.CD 燒錄軟體	3.網路通訊設備 4.磁帶機
後續 作業	1.建立普查母體資料檔	1.光學字元閱讀辨識系統 2.大型資料庫 (ORACLE) 3.瀏覽器 4.SAS、COBLE 5.磁帶館儲存管理系統 6.FILE-AID/ABEND-AID/STROBE	1.NT 伺服器 2.工作站 3.DVD 光碟櫃 4.磁帶櫃、磁帶館 5.網路通訊設備 6.IBM S390 主機
	2.專題分析	1.SAS、COBLE 2.FILE-AID/ABEND-AID/STROBE 3.WORD、EXCEL、漢書 4.VISIO	1.IBM S390 主機 2.磁帶館 3.個人電腦

參、本次普查資訊技術之主要興革

相較於歷次普查在資訊技術上的主要興革，其一為建置 OCR 系統處理普查大量資料的輸入，其二為連結多種公務登記檔案產生普查資訊及檢核普查相關資料，其三為運用 GIS 技術在直轄市建構數值圖檔運用，現分述於下：

一、OCR 系統之建置與應用

(一) OCR 系統概述

台灣 OCR 技術始於 1980 年代工研院的研發，目前對於手寫字母及數字的辨識率超過 95% 以上，而一般手寫文字若能配合字辭後處理輔助，則有尚可的辨識結果。其核心辨識技術係結合統計式與結構式文字特徵，再由高效率特徵轉換模組降低特徵維次，縮小儲存空間，提高辨識效率，最後以競爭性神經網路與字元型態資料庫進行文字辨識。其比對參用的字元型態資料

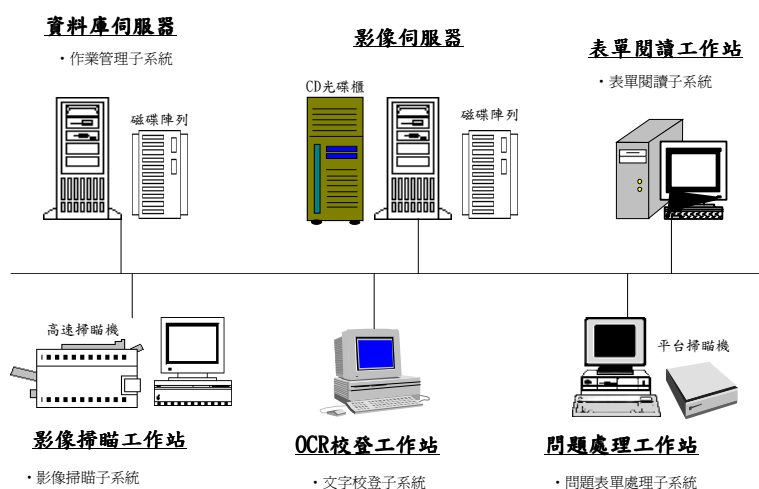
庫，係經長期蒐集累積國人手寫文字型態，並經辨識模組適當訓練而建立，未能辨識之字元轉由人工校登，技術上採行較佳的字元、欄位、整表等三階段校登方式，並先經資料庫技術排序處理後顯示，以避免畫面紊亂造成人員誤判，最後仍須進行欄位或整表校登的情形非常稀少。

(二) OCR 系統之建置

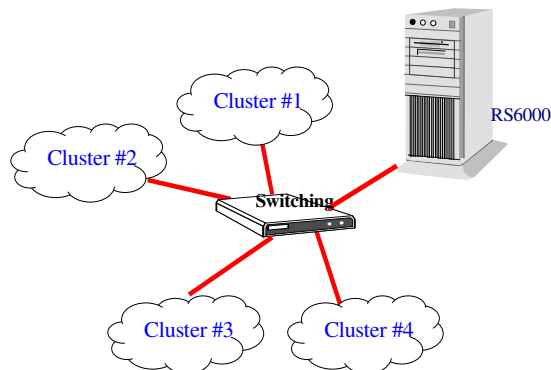
系統雖然委外進行開發，但本處人員以使用者身分提供意見亦積極參與作業過程，以使功能符合需求，設備得以延續運用。先期是以較小規模的人力資源調查為對象，將調查表重新按 OCR 所需版面設計，進行約一年的資料辨識及系統雛型測試作業，評估效果良好後，才展開複雜的普查 OCR 系統建置作業。

OCR 系統採主從架構設計，網路為 100Mbps Ethernet 架構，使用 Windows NT 作業系統及 Oracle 資料庫，因為高速掃描影像流量大，為顧及整體網路效能，乃規劃四個獨立處理單元 (Cluster) (圖一) 運作，並以處理單元管理主機 (RS6000) 經網路連結作為各處理單元間之溝通橋樑(圖二)，至於各單元內依系統功能可分為：「影像掃描子系統」、「影像瀏覽子系統」、「文字辨識子系統」、「文字校登子系統」、「問題表單處理子系統」、「資料轉換子系統」、「光碟製作子系統」、「影像查詢子系統」及「作業管理子系統」等九部分，設計上考慮掃描、辨識、線上校登、影像存取及光碟製作等流程的自動順暢，並兼顧作業管理上分派、監控及統計的需求。

圖一 處理單元 (Cluster) 示意圖



圖二 OCR 系統整體架構示意圖



(三) 應用結果與發展

本次普查表件計約 1200 萬張 (A3 大小)，OCR 系統平均每分鐘掃描 50 張，每 3 秒辨識一頁，全部掃描辨識工作於六十個工作天完成；運用約 110 人進行線上檢誤相關工作，結合影像調閱功能免除列印錯誤報表，合計五個月完成作業。總計字元校登率 1.21%，檢誤錯誤率 2.37%。而普查表以光碟影像儲存，不僅利於備份保存，更可節省五、六百坪置放空間。另有擷取普查表影像中文地址部分並予辨識建檔，而首次建立戶宅普查中文母體名冊資料供抽樣應用。至於系統相關設備與經驗則賡續為農漁、工商普查運用，不僅資源共享且作業亦更順暢。與上次普查資訊處理作業之比較見表二。

表二 近兩次普查資訊處理作業比較

項目	2000 年	1990 年
作業方法	OCR 作業	OMR 作業
時間	約五個月	約八個月
人力	110 人/月	180 人/月
正確率	99%以上	99%以上
中文地址辨識	可	不可
線上檢核	可	不可
調查表	紙張厚度 100 磅以上， 1200 萬張，A3 大小	須 OMR 專用紙， 700 萬張，B4 大小
套列受查戶資料	可	不可
光碟影像管理	可	不可
文件備份	可	不可
標準化	採用開放標準化之影像與資料庫， 易於整合其他應用軟體	無
WWW 查詢方式	可開放多點多人查詢	不可
列印錯誤報表	無	1300 箱
作業用磁帶	無	2000 捲

在這大規模的系統發展過程中，雖然曾遇到人員經驗的銜接、硬體設備與作業環境的不適、掃描辨識效果起伏現象、系統存取效能不佳等問題，但多能於短期內找出原因而予改善或解決。整體而言，OCR 系統對於大量資料的輸入處理作業，確實在時效、經費及品質上有很好的助益。未來將就通用性表單及整合自動註號功能發展，以充分發揮 OCR 系統在資料處理上所扮演的角色。

二、普查資料與公務檔案之連結運用

(一) 概述

面對調查環境的日益惡化，而公務資料電腦化的日益廣泛，善用公務登記資料已是潮流趨勢，就最早連結運用的北歐諸國，確有降低成本，增進統計資料品質之效果。本次普查規劃，即以公務檔案連結資料取代部分面訪問項，或是作為編製底冊及資料檢核之運用，例如以持有身心障礙手冊公務檔資料取代身心障礙狀況，以房屋稅籍檔資料取代建築類型、住宅完工年份及樓地板面積等；而連結戶籍檔檢核人口遺漏及性別、年齡等基本資料的確度。連結方法上，人口部分是以身分證統一編號為主鍵（Key）連結，另輔以性別、年齡、教育程度等屬性檢核處理主鍵連結不完整者；住宅部分則以地址為主鍵連結，另輔以身分證統一編號，以解決部分檔案間地址編碼不一致情形。至於調查資料與公務檔資料的定義、分類與內容範圍存有不一致者，如違建或公設面積、建築類型等，則做適當的調併或剔除。最後對於連結不成功者以熱檔法進行資料差補。

(二) 連結作業情形

為評估瞭解房屋稅籍檔連結效果，普查前分兩階段測試效果，第一階段與台北市戶籍檔連結測試；第二階段以南門里試查資料連結房屋稅籍檔，測試資料差異性。前階段初步連結成功率 87.78%，未連結成功原因，主要為房屋稅籍檔多址屬同一持有人但併為一筆陳示（約 3%）、地址陳示方式不一致、地址本身有錯、住宅已拆除及其他無法判識者等；至於不同特徵分類之未連結成功率相近，惟地區別間未成功率差異較大，其原因可能係房屋稅籍檔未隨區域重劃或門牌整編、更新地址資料所致。後階段就實際派員調查所得資料連結，其成功率約為九成，另於剔除非住家後比較面積，兩者間約有 2%之差異，整體運用上大致符合所需，而其正確性將因人為誤差之減少而提高。

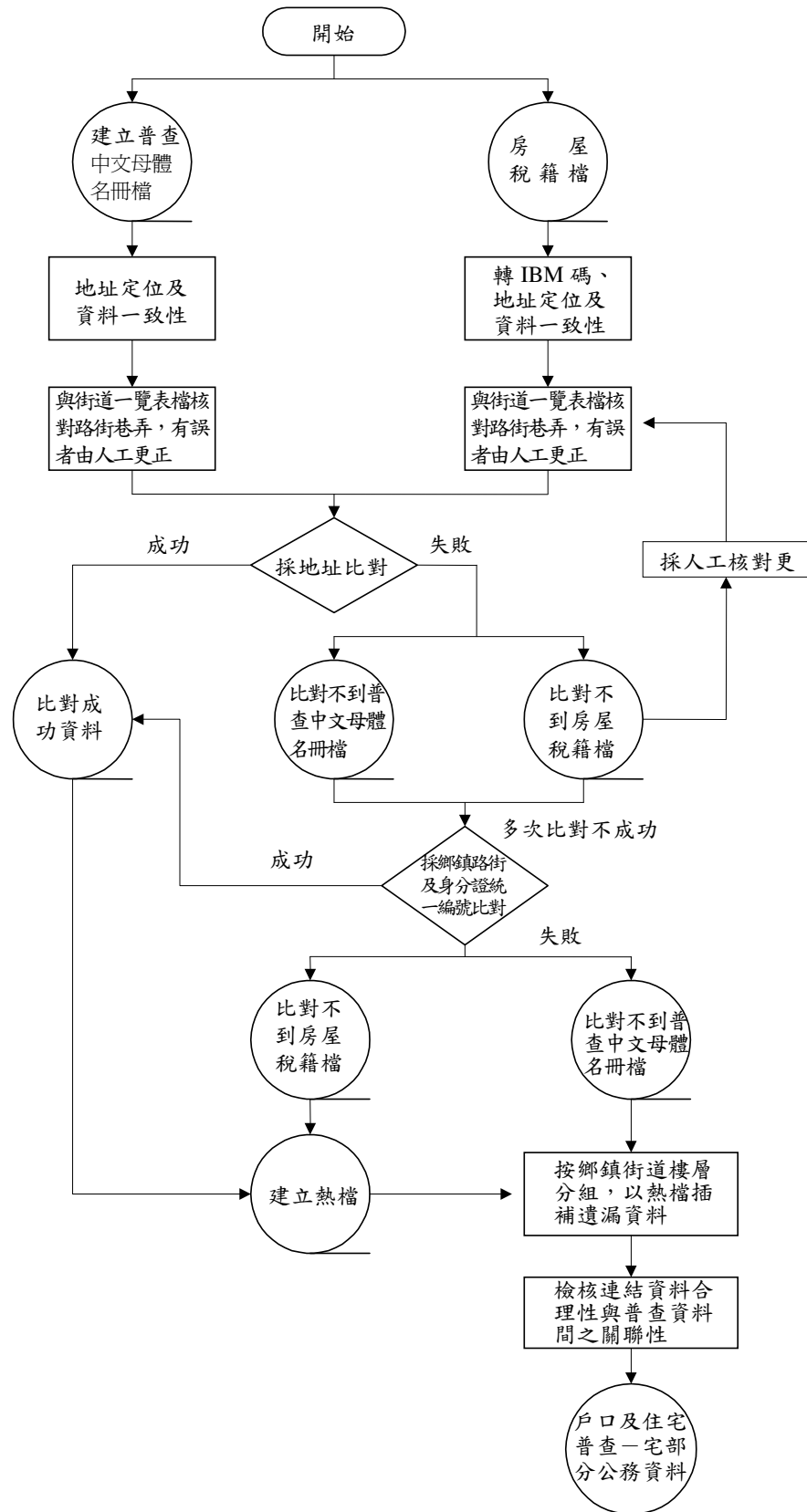
公務檔案的運用，由普查前的劃分普查區、編製普查底冊，到處理資料時的輔助檢核，以及連結產生替代問項之資料等，雖然取用的戶籍檔、房屋稅籍檔、身心障礙檔、村里門牌檔等來源不同，以致在資料格式或編碼、中文內碼、資料品質等常須做個別不同的加工處理或轉換，因而造成須用大量人工介入作業，缺乏良好的自動化作業，但在人員共同的努力下，仍逐一克服困難完成作業。實際普查檔資料與房屋稅籍檔連結流程如圖三。

(三) 連結運用結果與發展

整體連結運用結果，總計補列人口 0.55%，調整教育程度資料 3.25%，更正原住民身分 0.4%；而住宅連結成功率為 89.57%，對未連結者按鄉鎮市區、街道路名及樓層分層，採熟檔法隨機插補，各項特徵結構插補前後並無太大差異，整體效果尚屬可以。

公務檔案的連結運用困難關鍵在於資料品質及內涵或定義的差異，惟各公務檔案之建立各有其目的，歷史背景及作業嚴謹程度不同，因而常會影響作業時效及連結效果。有關公務檔案之連結運用已列入未來五年統計中程發展計畫中，將測試更多的公務登記資料，以適當取代調查資料，惟因各項資料存在前述運用上的問題，常非本處權責所能獨立解決，故未來發展的困難，不在於資訊技術上之不足，而是如何與各權責機關溝通協調以有效解決問題。

圖三 2000年戶口及住宅普查檔與房屋稅籍檔連結流程图



三、GIS 技術之運用

(一) 概述

普查是一種空間涵蓋面完整的調查，其優勢在於可以提供小地區與細分類之統計運用，惟調查之個體資料（Microdata）依法應予保密，因此適當的地理分類機制益顯重要。歷次普查是以”鄰”為普查區劃分基礎僅為分配工作量，而以”村里”作為普查最小統計地理單元，由於里鄰範圍易隨社經環境變動而被調整，且人口數量與特徵值亦差異大，因而減損普查資料整合及時間序列運用。若能藉由處理空間資料佳的 GIS 技術結合數質化地理圖檔，以明顯地物及自然地理特徵劃分普查區，將可建立範圍完整、適當、穩定且可長遠運用的普查地理單元，供為普查員責任區及普查資料整合之用。

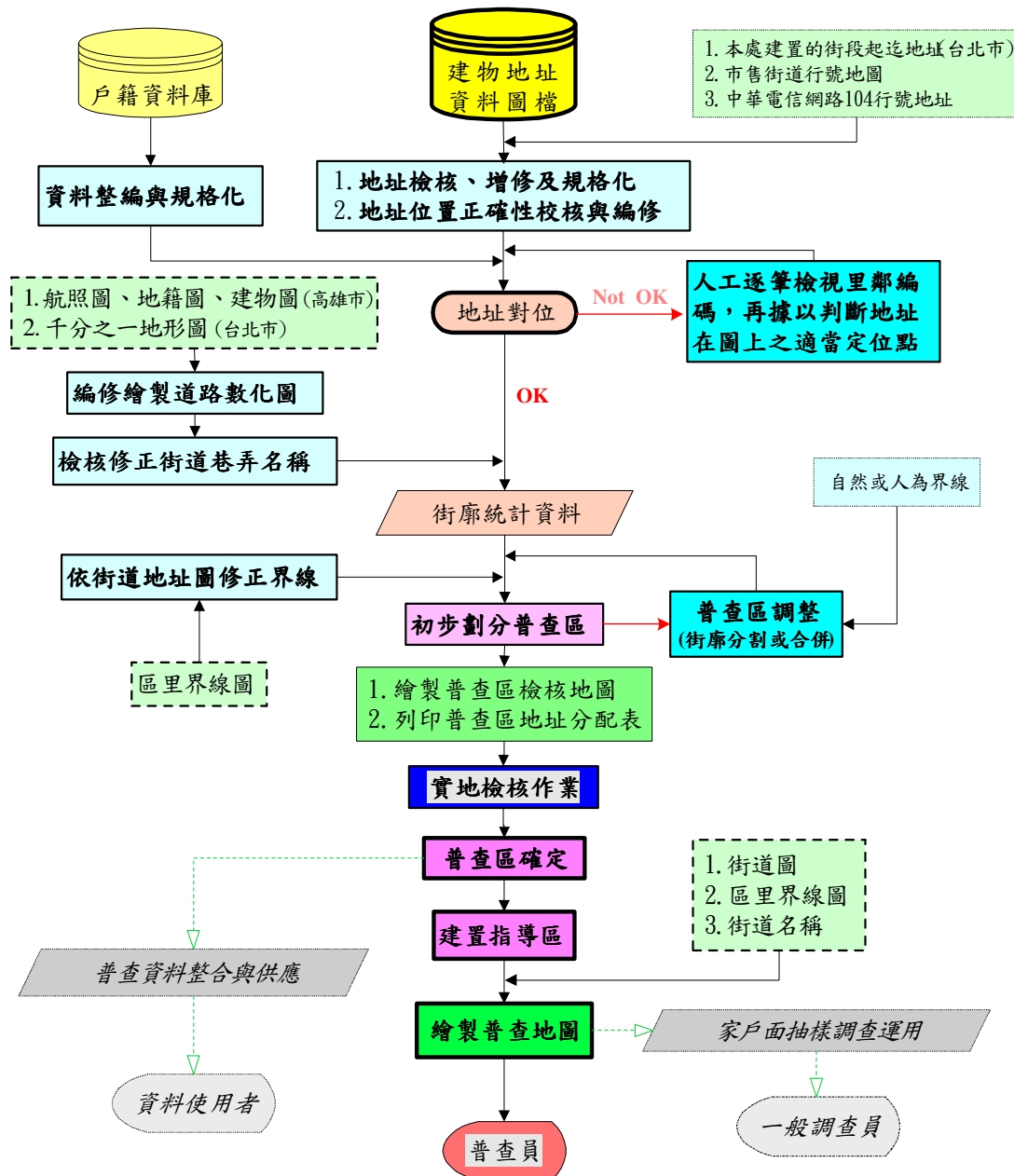
近年來隨內政部長期推動「國土資訊系統」的成果，取得台北及高雄兩直轄市地形圖及地址位置等數值圖檔，經適當的 GIS 及資料庫技術處理後，建立普查區、指導區圖層及其相關屬性統計，供實地訪查及資料整合連結運用。

(二) 普查區之建構

普查區是以 110±30 戶為劃分原則。屬性資料與普查區空間對應的方法，採用街段別（Street-Segment-Referencing）的地址對位（Address-Matching）方式連結，先建立各街段起迄門號，再以內插法連結各地址所應屬之普查區。

普查區劃定方法及過程如圖四，係將建物地址圖、地形圖、街道圖、行政界線圖等數值圖檔，利用 GIS 技術套疊與編修，再用地址對位方式連結戶籍資料，統計各街段戶數並予彙總為街廓（block）戶數。普查區即以街廓為基礎，由人工按劃分原則及地理環境等做適當的合併或分割，為期界線長遠的穩定明確，兩市均有約 2.5%（台北市 186 個、高雄市 103 個）普查區界線跨越里界。為確保普查區界線範圍及所含地址之適切完整，在普查前半年委由各里幹事進行實地檢核修正，結果約有超過八分之一的普查區須再予程度不等之編修，最後完成為期十四個月的台北市 7535 個、高雄市 4209 個普查區劃定作業。隨後並以 Avenue Script 語言撰寫自動繪製程式，利用既有雷射列印設備進行為期一個月的普查地圖（圖五）印製，提供普查員、指導員責任範圍地圖運用。

圖四 建置作業流程

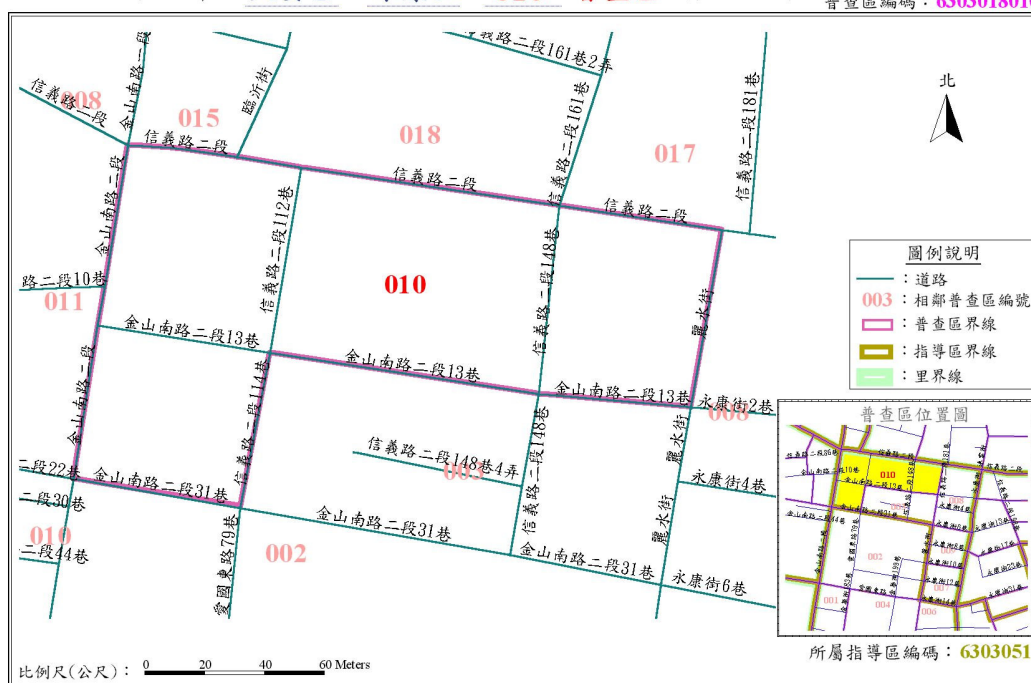


圖檔作業係在個人電腦環境中使用 PC ARC/INFO 及 Arcview 軟體進行處理，完成之普查區圖檔格式為 shape file，而戶籍資料等各種資料庫檔則經適當格式轉換後，利用 Access 資料庫軟體處理連結須用之屬性資料。各種圖檔除行政界線圖比例尺為 1:25000 外，其餘均為大比例尺 1:1000。由於原始地址及圖檔資料品質仍有瑕疵，且有 2~5 年的時間落差，因此花相當多的時間與方法在地址屬性資料及地圖位置的檢核與修正，以及街道圖重繪作業上。

圖五 普查地圖範例

台北市 大安區 永康里 010 普查區 普查地圖

普查區編碼：630318010



(三) 普查運用結果與發展

首次以 GIS 劃分普查區，避免責任區重疊或遺漏，運用相當順利成功。電腦自動化印製普查地圖，除可有效提升品質，較七十九年大幅節省六分之五的繪製經費。而以普查區為基本單元之普查地理資訊，亦可充實統計地區分類反映都市化發展差異、強化地理變數增進解析資料能力、供應小地區統計促進普查資料應用、展現統計地圖提供視覺化的空間圖像思考、作為統計地理資訊系統展現單元之基礎、供為家戶面調查之抽樣設計基礎等，可增廣應用範疇，提升普查整體效益。

雖然在緊湊的時程中順利完成作業，但所遭遇的困難，如缺乏標準化的地址編碼、各地區圖檔資源共享機制的不足、圖檔資料品質的差異、以及人員設備的不足等，均有待未來持續發展時應予重視者。全面性普查地理基礎的發展已列入未來五年統計中程發展計畫中，將按人口稠密及數值地圖的完備情形，從都市地區、城鎮地區、鄉村地區等順序逐步發展，不僅建構數值化普查區，亦將開發普查地理資訊系統，以輔助普查規劃及普查資料供應運用。

肆、結論與建議

一、結論

本次普查歷時五年的規劃、執行、編布資料等作業，較歷次普查運用更多的電腦方法與技

術，雖然發展過程中在方法、資料、人員配合、既有設備等方面，曾遭遇到程度不等的困難，但終能逐一克服，總結成效主要有：

- (一) 資料處理作業時間五個月，時程縮短；減省表件調閱人力，免除列印錯誤報表；節省存放空間，並可供光碟影像查詢調閱。
- (二) 農漁、工商普查及其他專案調查賡續運用 OCR 系統，資源充分利用。並擷取辨識普查表影像中文地址，據以首次建立普查中文母體名冊，供為各種抽樣應用。
- (三) 連結公務登記檔案，編製普查底冊減少漏查，修補普查資料增進品質。以房屋稅籍檔資料取代房屋面積、樓層、竣工年份、構造別等問項，節省調查經費與人力。
- (四) 導入 GIS 作業，建置北、高兩直轄市數值化普查區，繪製普查地圖節省製圖經費，供應小地區統計，增廣普查資料應用範疇。
- (五) 建立網際網路郵件收送系統及普查網頁，簡省文書作業，加速訊息傳遞效率，便利普查資訊取得。
- (六) 開發行政電子作業系統，使相關作業程序標準化，便於管理及減少錯誤，減輕人力負擔。

二、建議

- (一) 推展資訊系統的建置與整合，以充分發揮各項運用之效能。

目前本處正加強利用網際網路，研發 OCR 系統共通性表單及自動註號功能，擴大連結各機關公務登記資料，以及精進地理空間資料的整合運用等，若能建置適當的資訊系統，整合資料與流程，將更加發揮整體效能。

- (二) 標準化中文內碼及地址編碼，以促進資料的流通與連結運用。

各機關所建立電子檔，中文內碼未能一致，各地區地址編碼常缺乏規則性，致使資訊連結運用困難。為予改善，目前行政院主計處已在「第四期統計中程計畫」中，增進各機關電子檔案之標準化，期增進運用層面。

- (三) 建立政府登記資料的流通與共享機制，以使資源得到最大效用。

各政府機關公務登記電腦檔，在資料的內涵定義、涵蓋完整性、品質落差、格式標準化、登記時間差等仍有諸多問題尚待克服，須透過行政體系協調解決。而為避免資料索費限制公部門取用，宜建立各機關資料的流通及共享機制，以使各項資料獲得最大效用。

(四) 掌握設備環境的改善及技術趨勢，以能配合發展之需要。

運用新的方法及技術，常須引進新的軟硬體設備，或是就現有環境加以改善，因此適時適所的配合應予重視。對於資訊發展的趨向應多了解，以適當導入技術方法。

(五) 加強人員的資訊涵養，因應技術的不斷精進。

隨資訊技術多元發展且規模日趨複雜，亟需長期的規劃及維護管理能力人員，因此應積極加強人員資訊知能之培育，以延續並創建更多的運用成果。

三、結語

資訊技術在普查運用上所面臨的問題相當廣泛，其中有關規劃作業的改進、技術運用的加強、資訊系統的發展、資料的推廣運用等，可由本處自行逐步改善解決。惟在許多需用資料的取得流通及其內涵品質上，則因各有權責機關，有賴相關單位共同的協調配合，方可有效解決問題。我們除應加強與相關機關的聯繫協調，但為期事半功倍，在政府大力推動國家資訊通信基本建設（NII）的環境下，若能從資料源頭，特別是資料龐大、內涵豐富且具動態性質之人口戶籍、全民健保、軍公勞保、地籍建物及各種普查檔案，建立可完整結合的共享與共通機制，則在電子化政府（e-Government）逐漸形成之今日，將使各種檔案資料在資訊技術輔助下之運用更有效率、更具彈性，且大幅提升其運用範疇與效益。

參考文獻

- 1.行政院戶口普查處（1992）：“中華民國七十九年台閩地區戶口及住宅普查報告”
- 2.律基科技股份有限公司（2000）：“經濟部業界開發產業技術計畫書群集式泛用表單處理系統”
- 3.鄭敏祿、黃于玲、詹士賢（2000）：“運用公務檔案提升戶口及住宅普查效益”，八十九年統計學術研討會暨第九屆南區統計研討會
- 4.羅國華（2001）：“運用地理資訊技術建構普查區之實作與發展”，主計月刊第 550 期
- 5.Dekker, Arij L.（1994）：“Computer Methods in Population Census Data Processing”，International Statistical Review, Vol.62, No.1, pp.55-70.
- 6.United Nations（1996）：“Principles and Recommendations for Population and Housing Censuses”，United Nation published.