

第七章 結論與建議

以下分別針對婚育資料庫之維護與使用的設計、資料庫的建置階段、與各建置階段所可能需要的預算，提出說明和建議。

一、 婚育資料庫建置、維護、更新、與使用的電腦系統問題

婚育資料庫要保持其生命，必須能方便更新。這包括兩個面向：第一，新資料的定時登錄；第二，資料變項的增補、刪減。隨著社會經濟的演變，各個統計變項的重要性或許有所增減，有些過去被忽略的問題變得重要了，必須增建統計項目；有些原有的變項可能定義不恰當、或資料不準確，而必須淘汰。針對這個問題，在建置資料庫、編寫電腦程式之時，必須建立一般的通則，使維護者能夠以簡易的方式不斷的更新資料內容。考慮到這個問題，在建資料庫時，電腦系統的設計有兩個方向。一種是比較傳統的作法：資料蒐集者事先製作千百個次數分配表（frequency tables）、交叉表或其他加總整理過的表，然後將這些變項的數據資料放在 Excel 檔裡（文字資料，例如法規的資料，放在 Word 檔裡），程式設計者再撰寫配套的程式，整批讀取這些資料，送進資料庫的電腦環境之中。為了達到這個目的，程式設計者必須先確定數位化物件的「命名架構」，以便資料庫能不斷地增錄往後各期的資料。此外，在考慮「數據資料」（統計資料）與「名稱資料」（變項名稱）在資料庫中的連結時，不單要指明檔案的「實體位置」（physical location），也要指明其「邏輯位置」（logical location），以便資料的延伸、擴充、和儲存位置的更動。而整體的資料存取環境必須是「關聯式資料庫」（例如 Oracle, MySQL, Postgres SQL 等），以便將數據資料「正規化」，使維護者能根據使用者之需要來改變資料的結構。這種資料庫環境的設計也可以讓維護者以單筆

資料一一輸入的方式來更新資料庫。對於資料庫的使用者而言，上述的資料庫電腦系統與我們在第二節的資料查詢範例所表現的類似，可以讓使用者自由組合已經整理好的次級資料。同時，由於各變項的統計表已經以人工的方式事先製作好，使用者也可以直接進入婚育資料庫首頁的「資料名詞字典」（或「資料目錄」、「資料索引」）選項，然後直接點選期中的各名詞（或項目），直接取得數據。但是，這種傳統的資料庫電腦系統無法像「資料探勘」（data mining）的作法一樣，讓使用者利用資料庫中的資料，以和資料互動的方式，在線上直接計算出某些自己想要的統計量。

可做「資料探勘」（data mining）或互動式線上計算的資料庫電腦系統有好幾種，其一是 Data Documentation Initiative（DDI）。DDI 是符合 Extensible Markup Language（XML）的網路資料庫規格，1998 年才問世，已被美國大型的人口資料庫（包括人口普查資料庫）採用為儲存資料的規格。這種資料庫的規格使資料庫的維護更新更方便，資料內容的存取都由程式操作填入，不必經過人工的複製和貼入；網頁的畫面也不易受到資料增補更新的影響。同時，它有「資料探勘」（data mining）的功能，使用者可以在查詢資料之後直接在線上運算出一些統計量。如果我們要將原始資料（初級資料）放到婚育資料庫之中，則 DDI 的規格比前述的傳統資料庫電腦系統要恰當。將查詢者的電腦直接連上原始資料，便可利用線上運算的程式，依據原始資料直接算出各種統計指標，而不必像前面那一種傳統的作法，事先製作千百個次數分配表（frequency tables）、交叉表或其他加總整理過的表，然後放在資料庫裡面。線上運算的作法有利於指標資料的即時更新，和分配次數（frequency）計算、交叉分析等基本的統計作業。

另一種可做互動式線上計算的、與 DDI 類似但功能比較簡單的

資料庫電腦系統是「線上即時分析」(Online Analytical Processing, OLAP)。目前, OLAP 在台灣的使用比較普遍, 也是行政院衛生署「國民健康局」所使用的資料庫電腦系統。OLAP 是 E. F. Codd 在 1993 年提出的一種資料分析、數據分析系統, 為「連線交易處理系統」(Online Transaction Processing, OLTP)之改良。OLTP 是傳統的關聯式資料庫, 處理日常的事務如銀行交易等。OLAP 則是「資料倉庫系統」(data warehouse)的應用, 支持多維環境下的查詢和統計工作, 並提供直觀易懂的查詢結果。OLAP 利用數學和物理上的「維」(dimensions)的概念, 來分析處理資料。「維」是我們觀察世界的角度, 代表層次的關係。我們可把一個對象的多種屬性定義為多個「維」, 使用者對不同維上的數據做比較。因此, OLAP 可以說是多維資料分析工具的集合, 對資料進行查詢、分析、和統計。例如, 我們在觀察人們的婚育行為時, 可從其居住地區、性別、年齡、教育程度、族群等不同的角度來觀察各屬性之人的表現與其差異。這裡的居住地區、性別、年齡、教育程度、族群等等就是「維」。而這些維的各種組合和度量指標是 OLAP 的分析對象, OLAP 對以多維形式組織

表 7-1 OLAP 資料庫電腦系統之特點

	OLTP	OLAP
用戶	一般作業人員	決策人員、管理人員
功能	日常操作處理	分析決策
數據資料	當前的、最新的、細節的, 二維的、分立的	歷史的、聚集的、多維的、集合的
存取	讀/寫數十條記錄	讀取上百萬條紀錄
工作	簡單的事務	複雜的查詢與計算
用戶數	上千個	上百個
Database 之大小	100MB-GB	100GB-TB

資料來源：<http://www.pku-ht.com/>

起來的數據採取切片（Slice）、切塊（Dice）、鑽取（Drill-down 和 Roll-up）、旋轉（Pivot）等各種分析動作，以求剖析數據。使用者能從多個角度萃取資料庫中的數據，從而掌握包含在數據中的訊息（參考 <http://www.pku-ht.com/>）。表 7-1 列出 OLAP 的特點。

與 DDI 類似，OLAP 規格的系統使資料庫的建立和維護比傳統的資料庫電腦系統要方便，不必經過人工的複製和貼入；網頁的畫面也不易受到資料增補更新的影響。同時，將查詢者的電腦直接連上原始資料，使用者也可以利用 OLAP 線上運算的程式，針對數據資料直接算出一些基本的統計指標，而不必像傳統的作法，事先製作千百個次數分配表、交叉表或其他加總整理過的統計表。

二、 婚育資料庫建置的各階段進程建議與預算粗估

我們的研究顯示，符合理想的婚育資料庫之建置工程可能頗為龐大，有原始資料與次級資料的問題、有國外資料蒐集的問題、還有資料庫電腦系統的問題。而資料庫電腦系統的選擇，又與婚育資料庫是不是要放進原始的資料，以直接進行線上運算有關。因此，婚育資料庫能夠做到何種程度，與技術、人力、預算、和法律限制都有關。以下將資料庫的建置工作，依其工作量的大小和難易程度，分成五個年度，同時粗估各年工作的潛在花費。

1. 第一年

第一年工作的內涵，是整合性的婚育統計資料庫的基礎工作，建立其基本的型態。其關鍵工作有二：一是基本資料之蒐集，二是資料庫電腦系統的選擇、建置、和試跑。就資料蒐集而言，我們建議，在第一年的工作中，僅先蒐集統計資料，法規資料暫不收錄。而統計資料庫也僅先蒐集國內的次級資料，無次級資料之形式的原始資料集不

錄，國外的婚育統計資料也暫時不收錄。其詳如下：

- (1) 國內的婚育統計資料：只蒐集次級資料，將其建置於資料庫電腦系統之中。
- (2) 開始建立「資料目錄」（「資料名詞字典」或「資料索引」），置於資料庫電腦系統之中：「資料目錄」的項目在第六章第三節已初步列出，正式建置時須中英文並列，並說明各項（各名詞）之定義、資料來源（哪一種公務資料、哪一種調查資料）。如第六章所述，此一目錄除了按第六章第三節的方式依照大約的類別排列之外，另分別按中文筆劃和英文字母排列。此一工作將耗費不少人力和時間，查詢者閱覽起來也頗費時。因此，我們建議，此前先建一個簡明版的「資料名詞字典」，如表 7-2 所示，同樣建入資料庫電腦系統的首頁之中。
- (3) 國外的婚育統計資料：暫不蒐羅。
- (4) 國內的婚育法規與政策資料：暫不蒐羅。
- (5) 國外的婚育法規與政策資料：暫不蒐羅。
- (6) 資料庫電腦系統：就下列三者擇一，(a) 傳統的資料庫電腦系統，(b) 「線上即時分析」(OLAP)，(c) Data Documentation Initiative (DDI)。如果將來要把原始資料建入婚育資料庫，並且讓使用者能夠直接對原始資料做線上運算，則從第一年開始，就不能用傳統的資料庫電腦系統，而得用 OLAP 或 DDI 等有線上運算之功能者。我們建議採用「線上即時分析」(OLAP) 系統，如此則資料庫的電腦

系統既可讀取、組合次級資料，也可以分析未來可能建入的原始資料集。

- (7) 建置「資料庫導覽」手冊，簡單地說明資料庫資料庫的主要內容、架構（將圖 6-1 和圖 6-2 的內容放入）、各種資料的來源機構等等，置於資料庫網站的首頁。
- (8) 設定連結的網站，並在使用者嘗試進入被連結的網站之時先介紹該網站的主要內容與資訊，因此需要編寫各連結網站的簡介。
- (9) 編寫資料庫更新及擴充的「標準作業程序」（SOP）和網站連結之版權問題的 SOP，供資料庫的建置、操作、與維護單位之參考。
- (10) 第一年的建置預算粗估：(a) 若採用傳統的資料庫電腦系統，第一階段規模的建置費用可能需 400 萬元以上，包括：資料庫電腦系統（資料分析設計、流程設計、正規化工作、資料輸入與管理工具）之建置約 50 萬元；其餘為以人工方式建置資料目錄、事先製作次數分配表、進行資料之人工複製、工作管理、資料內容和資料格式之正確度檢查等等的費用。(b) 若用「線上即時分析」(OLAP)，系統建置費亦約 50 萬元；人工方式建置資料目錄的花費也相當，但事先製作次數分配表、進行資料之人工複製、資料內容和資料格式之正確度檢查等等的工作減省很多。預估總花費約 250 萬元左右。(c) 若用 DDI 技術，由於其目前在台灣的應用尚不廣，純就電腦系統本身的資本與技術費用而言，可能較貴；但與其搭配的非電腦系統或非程式相關的人力

費用，不見得更貴。

表 7-2 人口資料重要名詞清單

基本人口資料	生育資料	死亡資料	婚姻資料
靜態人口資料	總出生數	總死亡數	婚姻狀況組成
動態人口資料	年齡別/性別/出生數	年齡別/性別死亡數	未婚人數
生命統計	婚姻狀況別出生數	粗死亡率	結婚人數
簡易生命表	粗生育率	年齡別/性別死亡率	結婚對數
完整生命表	育齡婦女一般生育率	嬰幼兒死亡率	結婚年齡
總人口數	育齡婦女年齡別生育率	死產率	離婚人數
年底人口數	有偶婦女年齡別生育率	死因別死亡率	粗結婚率
年中人口數	時期生育率	預期壽命（平均餘命）	粗離婚率
人口金字塔	年輪生育率		平均初婚年齡
年齡結構	粗繁殖率		結婚平均年齡
三階段年齡人口	淨繁殖率		初婚數
人口的平均年齡	替換水準		初婚率
性別結構	胎次別生育數		離婚數
教育組成	胎次別生育率		離婚率
基礎年	生育步調		喪偶數
基準人口	期望子女數		喪偶率
年輪（世代）	出養子女數		離婚再婚數
年輪率	領養子女數		離婚再婚率
人口成長率	生育間隔		喪偶再婚數
自然增加率	墮胎數		喪偶再婚率
社會增加率	墮胎率		年齡別結婚率
淨遷移數	非婚生數		年齡別離婚率
移入人口數	非婚生率		年齡別喪偶率
移出人口數			婚姻狀態移轉率
人口密度			同居人數
依賴比			可婚人數
扶養比			
性別比			
戶口數			
戶長率			
子女離家率			
家庭生命週期			

2. 第二年

我們建議，第二年工作仍然僅蒐集次級統計資料，原始的統計資

料和法規資料暫不收錄。工作的重點是，開始蒐集國外的婚育統計資料（並將其英文文字的部分轉成中文）。其詳如下：

（1）國內的婚育統計資料：加強蒐集次級資料，根據專家或各方面之意見，補足第一年資料蒐集工作的不足之處。同時，本計畫書在前面提到，目前的某些婚育統計指標頗有偏誤和欠缺之處。因此，為使婚育資料庫更為有用，應補足這方面的缺憾。我們建議，配合婚育資料庫的擴充，內政部進行調查項目之增補。上述的同居資料、婚外生育的資料若在婚育資料庫建置的第一年尚未採集，建議配合第二年之擴充，開始調查蒐集。但這部分的調查，與婚育資料庫建置本身是不同的工作，我們不將其費用列在婚育資料庫的增建預算之內。若政府機構增加新的、重要的統計項目，並產生了資料（例如同居資料、婚外生育的資料），則應當將其原始檔案展換成次級資料，經專家審查確認之後，增建進入婚育統計資料庫中。

（2）國外的婚育統計資料：只蒐集次級資料；同時，資料的來源暫時集中於聯合國人口署的人口統計資料，歐盟人口婚育資料網站（EUROSTA），美國人口普查局的國際人口資料庫，聯合國歐洲經濟組織人口活動小組（UNECE/PAU）的婚育調查資料（Fertility and Family Surveys），世界銀行的兩性統計資料庫 GenderStats（Database of Gender Statistics），私人機構 Population Reference Bureau 等等。各國政府的人口統計網站雖然資料豐富，但逐一檢索、蒐集的工作量龐大，恐不是短期內能完成者，待第三年再開始蒐羅。

- (3) 國內的婚育法規與政策資料：暫不蒐羅。
- (4) 國外的婚育法規與政策資料：暫不蒐羅。
- (5) 資料庫電腦系統：在第一年就必須確定，我們已建議採用「線上即時分析」(OLAP)系統。
- (6) 研究文獻之蒐集：按照資料庫來源類別，將曾經利用這些資料來源完成論文者之資訊予以蒐集，並將其論文摘要列載於資料庫上。鑑於文獻數量可能不少，我們建議分階段來建立研究文獻的目錄，同時蒐羅的對象以利用調查統計者為主，凡利用公務統計者，主選出較有代表性的文章。本年度先篩選出較為經典的文章，並經專家審查確認，再提供使用者參考。較詳細的研究文獻目錄可以後續的三年為期，逐年增建。
- (7) 生育預警指標之研究：從本年度開始，可以開始研究、建立生育行為的預警指標，包括生育率本身、各種結婚和同居行為的數據、婦女勞動參與率、婦女的教育程度、總體失業率與兩性的失業率，等等。有了個別的潛在指標之後，可嘗試建立一個整合各指標為單一指標的加權公式。
- (8) 增修「資料目錄」、「資料庫導覽」手冊、資料庫更新及擴充的「標準作業程序」(SOP)、網站連結之版權問題的SOP，等等。
- (9) 第二年增建預算粗估（不含統計資料庫的日常維護和小規模更新的費用）：若用「線上即時分析」(OLAP)，系統建置費在第一年已支出；本年度的工作主要在蒐集、整理國

外婚育資料、研究文獻之蒐集、預警指標之研究。預估總花費約 200 萬元左右。

3. 第三年

第三年個年度加強國外婚育統計資料的蒐集，將蒐羅的來源擴張到各國政府的網站，同時開始蒐集國內外的法規資料的蒐集。其詳如下：

- (1) 國內的婚育統計資料：加強蒐集次級資料，建入政府機構新增的、重要的婚育統計項目。
- (2) 國外的婚育統計資料：只蒐集次級資料，但開始有系統地蒐羅各國政府的人口統計網站的資料，並建入資料庫中。選擇國家的標準有三：(a) 婚育資料豐富（例如美國），(b) 人口與婚育型態變遷經驗可供台灣參考者（例如日本、瑞典）(c) 與台灣處在類似的人口與婚育變遷階段者（例如香港、新加坡、韓國）。初步約選十個以內的國家或地區。
- (3) 國內的婚育法規與政策資料：普遍蒐集國內的婚育法規資料，但暫時不蒐集白皮書以外的政策資料。
- (4) 國外的婚育法規與政策資料：不逐一蒐集各國的法規資料，僅收國際規約、宣言、或某些代表性國家的代表性婚育法規。
- (5) 資料庫電腦系統：建議採用「線上即時分析」(OLAP) 系統。
- (6) 增修「資料目錄」、「資料庫導覽」手冊、資料庫更新及擴充的「標準作業程序」(SOP)、網站連結之版權問題的 SOP、

增建研究文獻之目錄、改進生育預警指標，等等。

- (7) 第三年之增建預算粗估（不含統計資料庫的日常維護和小規模更新的費用）：若用「線上即時分析」（OLAP），系統建置費在第一年已支出；本年度的工作主要在蒐集、整理各國政府的婚育資料，並蒐集國內外的婚育法規與政策資料，增強研究文獻目錄，預估總花費約 200 萬元左右。

4. 第四年

我們建議，第四個年度開始將某些原始資料集建入婚育統計資料庫，同時加強國內婚育政策資料的蒐集。其詳如下：

- (1) 國內的婚育統計資料：對次級資料做必要的增補、更新；並選擇一個到兩個沒有版權問題的原始資料集，建入婚育統計資料庫。我們建議在本年度中選擇調查統計（非公務統計）的原始資料集開始嘗試，例如衛生署的「家庭生育力調查」（KAP）等。
- (2) 國外的婚育統計資料：加強各國政府次級婚育統計資料之蒐集，可以增加二到三個國家的資料。
- (3) 國內的婚育法規與政策資料：開始普遍蒐集白皮書以外的政策資料。
- (4) 國外的婚育法規與政策資料：同前，僅選有特殊代表性者。
- (5) 資料庫電腦系統：建議採用「線上即時分析」（OLAP）系統。
- (6) 增修「資料目錄」、「資料庫導覽」手冊、資料庫更新及擴

充的「標準作業程序」(SOP)、網站連結之版權問題的 SOP、增建研究文獻之目錄、改進生育預警指標，等等。

- (7) 第四年之增建預算粗估(不含統計資料庫的日常維護和小規模更新的費用):本年度的主要工作是，建入原始資料集並測試其線上分析的功能，增加二到三個國家的婚育統計資料，並蒐集白皮書以外的國內婚育政策資料。預估總花費約 200 萬元左右。

5. 第五年

第五年的工作主要在於：對前四年的工作成果改進、補強，並擴充原始資料集，甚至考慮納入某些公務統計的原始資料。

- (1) 國內的婚育統計資料：加強次級資料之蒐集；擴充原始資料集，可嘗試納入某些公務統計，但是，不做因法律限制而無法做的原始資料集串接。
- (2) 國外的婚育統計資料：只蒐集次級資料，但可再增加一到二個國家的資料。
- (3) 國內的婚育法規與政策資料：做必要之增補、更新。
- (4) 國外的婚育法規與政策資料：做必要之增補、更新。
- (5) 資料庫電腦系統：建議採用「線上即時分析」(OLAP)系統。
- (6) 增修「資料目錄」、「資料庫導覽」手冊、資料庫更新及擴充的「標準作業程序」(SOP)、網站連結之版權問題的 SOP、增建研究文獻之目錄、改進生育預警指標，等等。

- (7) 第五年之增建預算粗估（不含統計資料庫的日常維護和小規模更新的費用）：本年度的主要工作是：擴充原始資料集，增加一到二個國家的婚育統計資料，並對婚育資料庫的操作功能做總檢視。預估總花費約 200 萬元左右。

表 7-3 婚育資料庫建置階段

	建置內容	主辦單位	建議之 協辦單位
第一年	①國內婚育資料：次級資料	內政部 戶政司	衛生署 主計處
	②將「資料目錄」(「資料名詞字典」) 建入資料庫中	內政部 戶政司	衛生署 主計處
	③國外的婚育統計資料：暫不蒐羅		
	④國內的婚育法規與政策資料：暫不蒐羅		
	⑤國外的婚育法規與政策資料：暫不蒐羅		
	⑥資料庫電腦系統：OLAP	內政部 戶政司	衛生署 主計處
	⑦建置「資料庫導覽」手冊	內政部 戶政司	衛生署 主計處
	⑧編寫資料庫更新及擴充的「標準作業程序」(SOP) 和網站連結之版權問題的 SOP	內政部 戶政司	衛生署 主計處
	⑨設定連結的網站，編寫各網站的簡介	內政部 戶政司	衛生署 主計處
	⑩第一年的建置預算粗估：250 萬元	內政部 戶政司	
第二年	①國內婚育資料：加強次級資料之蒐集	內政部 戶政司	衛生署 主計處
	②國外的婚育統計資料：蒐集國際組織網站的各國次級婚育資料	內政部 戶政司	衛生署 主計處
	③國內的婚育法規與政策資料：暫不蒐羅		
	④國外的婚育法規與政策資料：暫不蒐羅		
	⑤資料庫電腦系統：OLAP	內政部 戶政司	衛生署 主計處
	⑥建立婚育研究文獻之目錄(主要的蒐集對象為曾利用各婚育資料集之來源所做的研究)	內政部 戶政司	衛生署
	⑦嘗試建立生育預警指標	內政部 戶政司	衛生署
	⑧增修「資料目錄」、「資料庫導覽」手冊、資料庫更新及擴充的「標準作業程序」(SOP)、網站連結之版權問題的 SOP，等等。	內政部 戶政司	衛生署 主計處
	⑨第二年的建置預算粗估：200 萬元	內政部 戶政司	

	建置內容	主辦單位	建議之 協辦單位
第三年	①國內的婚育統計資料：加強次級資料之蒐集	內政部 戶政司	衛生署 主計處
	②國外的婚育統計資料：次級資料，開始各國政府的婚育統計資料之蒐集	內政部 戶政司	衛生署 主計處
	③國內法規與政策資料：開始蒐集國內的婚育法規資料，但暫不蒐集白皮書以外的政策資料	內政部 戶政司	衛生署 法務部
	④國外法規與政策資料：代表性規約	內政部 戶政司	衛生署 法務部
	⑤資料庫電腦系統：OLAP	內政部 戶政司	衛生署 主計處
	⑥增修「資料目錄」、「資料庫導覽」手冊、資料庫更新及擴充的「標準作業程序」(SOP)、網站連結之版權問題的 SOP、增建研究文獻之目錄、改進生育預警指標，等等。	內政部 戶政司	衛生署 主計處
	⑦第三年的建置預算粗估：200 萬元	內政部 戶政司	
第四年	①國內的婚育統計資料：加強次級資料之蒐集，開始將原始資料集建入資料庫系統	內政部 戶政司	衛生署 主計處
	②國外的婚育統計資料：次級資料，增加對各國政府的婚育統計資料之蒐集	內政部 戶政司	衛生署 主計處
	③國內法規與政策資料：開始蒐集白皮書以外的政策資料	內政部 戶政司	衛生署 法務部
	④國外法規與政策資料：代表性規約	內政部 戶政司	衛生署 法務部
	⑤資料庫電腦系統：OLAP	內政部 戶政司	衛生署 主計處
	⑥增修「資料目錄」、「資料庫導覽」手冊、資料庫更新及擴充的「標準作業程序」(SOP)、網站連結之版權問題的 SOP、增建研究文獻之目錄、改進生育預警指標，等等。	內政部 戶政司	衛生署 主計處
	⑦第四年的建置預算粗估：200 萬元	內政部 戶政司	
第五年	①國內的婚育統計資料：加強次級資料之蒐集，擴充原始資料集	內政部 戶政司	衛生署 主計處
	②國外的婚育統計資料：次級資料，再擴充資料之蒐集國家對象	內政部 戶政司	衛生署 主計處
	③國內法規與政策資料：做必要之增補、更新	內政部 戶政司	衛生署 法務部

	建置內容	主辦單位	建議之 協辦單位
	④國外法規與政策資料：做必要之增補、更新	內政部 戶政司	衛生署 法務部
	⑤資料庫電腦系統：OLAP	內政部 戶政司	衛生署 主計處
	⑥增修「資料目錄」、「資料庫導覽」手冊、資料庫更新及擴充的「標準作業程序」(SOP)、網站連結之版權問題的 SOP、增建研究文獻之目錄、改進生育預警指標，等等。	內政部 戶政司	衛生署 主計處
	⑦第五年的建置預算粗估：200 萬元	內政部 戶政司	