

附錄二 「國內外婚育資料庫建置研究計畫」專家座談會

壹、時間：95 年 9 月 25 日（星期一）上午 10 時整

貳、地點：台灣綜合研究院台北辦事處（台北市敦化南路一段二號四樓）

參、主席：李安妮（台灣綜合研究院副院長，「國內外婚育資料庫建制研究計畫」主持人）

記錄：陳思蓉

肆、出席單位及人員

林季平 中央研究院調查研究專題中心副研究員

徐 美 國立台北大學經濟系教授

莊義利 行政院衛生署人口與健康調查中心主任

許汶瑛 行政院主計處第四局第四科科长

陳信木 國立政治大學社會學系教授

陳敬宏 內政部統計處統計長

陳寬政 長庚大學醫務管理學系教授

楊文山 中央研究院社會學研究所研究員

蘇清朝 內政部戶政司副司長（因公不克出席）

陳子和 內政部戶政司

黃美正 內政部戶政司

楊靜利 南華大學社會學研究所副教授

陳 筆 台灣綜合研究院研二所副所長

伍、宣布開會 台灣綜合研究院簡報（略）

陸、討論

一、陳敬宏

1、初級資料與次級資料：

就公務統計而言，原始的個人戶籍登記資料是初級資料，發佈出來的統計結果叫次級資料。就調查統計而言，個別的調查紀錄是初級資料，發佈出來的統計結果叫次級資料。初級資料與次級資料皆應建入婚育資料庫中。

我認為把初級資料登記放進婚育資料庫裡是很重要的，當然，原始資料的個人 ID 必須拿掉，這是一定的。放入初級資料的好處是，使用者想算什麼，他可以根據原始資料自己去算，你不必事先替他想他可能要什麼，然後算出次級資料給他用，這樣比較沒有彈性。例如中研院的調查專題研究中心有各種原始資料（raw data），其中也有其他部會提供給它的。當然，也許不必用初級資料和次級資料的名詞，而以公務統計和調查統計代之。

婚育資料庫若沒原始資料，會有一個潛在問題，那就是日後若有新增的統計項目，如果原先未建有原始資料，則如何把新增的統計項目的資料內容追溯到過去？原始資料可容許最大彈性的運用。

另外，調查統計的原始資料可以產生很多的統計項目。如果婚育資料庫不放入原始資料（初級資料），則你必須花很大的功夫從調查統計的原始資料裡篩選項目。而你要選出哪些項目？如果婚育資料庫的使用者找不到某些他要的項目，那是原始調查本來就沒有，還是它有，但你建資料庫時沒有把它選進去？如果把原始資料放進資料庫，就可避免這種迷惑。

整合原始資料的工作也許太過龐大，但我建議規劃單位還是要把整合原始資料庫的方向放入資料庫規劃的架構之中。至於何時實施原始資料整合的工作，由內政部戶政司以後來考慮。

2、婚育資料庫的架構與內容

死亡與生命表的資料要不要放進婚育資料庫？死亡和生命表是人口預測的關鍵要素，所以，可在資料庫的第三層同時放入死亡、生命表和人口預測的大項。

關於法規資料庫的部份，台綜院的計劃書和期中報告提到公共政策對生育率的影響。事實上，目前新竹縣市的生育率高於其他地區，而新竹縣市的生育補助也比較高。因此，除了法規之外，也可以把政策資料放進去，而成為法規政策資料庫。政策的執行常又涉及經費支出，是不是也可把婚育相關政策的經費支出放進資料庫？

婚育資料庫未來要放入調查統計的資料，通常調查統計的涵蓋面很廣，在婚育的內容之外，也可能包括家庭、經濟、衛生、醫療等面相。因此，如何將調查統計的結果放進婚育資料庫，必須考慮，是將調查統計單獨劃分出來，而下面列婚育、經濟等各大項呢？還是先列出各大項，再將調查統計的結果依其類別列入各項之下？請考慮。

另外，目前有各種公私部門的調查統計，將來把這些調查統計的項目和統計結果納入資料庫時，應當經過恰當的審核統計品質與確認的過程。因此在規劃婚育資料庫時，也應該一併考慮資料審核機制的問題。

此外，目前婚育資料庫所建議納列的主要是既有的統計項目，而把它們整合起來。但是對目前沒有的、未來應當新增的項目，也應該

列舉出來，以備未來做資料蒐集。

最後對於婚育資料庫檢索時所用的分層變項提供一些意見：(1) 在地區別中，已有各縣市的資料，可以拿掉台灣地區的分層。(2) 在人口項，應加入國籍別的分層，在生育項也加入生母之國籍別的分層。(3) 生育項可加入嬰兒胎次別的分層。(4) 死亡項可加以加入死因別（例如與生育有關的原因而致死）的分層。(5) 社會經濟項可就勞動力、教育、醫療衛生、托育等類別來分層。

二、陳寬政

1、初級資料與次級資料：

初級資料可以容許許多有彈性而靈活的運用，例如把健保資料檔和身份證資料檔合併起來運用。照這作法來看，各個政府部門的許多原始資料檔都可以利用 ID 碼結合起來作靈活的使用，有許多分析上的可能性，這是初級資料的好處。

婚育資料庫是否放進原始資料或整合原始資料，涉及到計畫的目標是什麼。中研院調查中心的原始資料只是個檔案庫，資料放在那裡，你要用就自己用，它沒有替你整合起來，也沒有寫好程式讓你用它來計算加總資料。若要整合各部會的原始的資料庫，固然結果會很 powerful，但這還有程式的問題。這要花很多功夫，也有隱私安全的問題，我想這麼龐大的工作不是婚育資料庫的基本目標。這個資料庫的目的還是在於：把政府發佈的各種婚姻與生育的統計結果整合起來，方便使用。當然，現在網上看到的出生率、死亡率資料都是從戶籍資料來的。理想上來說，如果內政部能在其中央電腦系統上建一個程式，把電子化的戶籍資料利用 ID 和其他部會的原始資料（例如健

保資料)串在一起，在網上開放供人讀取使用，這是功能最大、最有彈性的(當然不讓使用者看到ID碼)。

2、生命表、死亡資料與疾病資料：

婚育資料庫的規劃之中列有生命表和死亡的資料，那麼疾病的資料是不是也要列入？疾病資料從健保檔中可以取得，不過，我並不是說婚育資料庫必須列入疾病資料，這只是給大家參考。

台灣人口的粗死亡率是千分之五，是世界最低的。但這是 17 個年齡組的死亡率加權平均的結果，依各年齡層人口的比重而定，生育率的計算也是。若要放進死亡的資料，則除了性別、年齡別、有偶、無偶的分層變項的控制之外，死因的資料也很重要。1971 年以來，衛生署和前省衛生處有死因的資料，但前衛生處的死因檔已經佚失。

3、生育資料：

在人口學中，生育率資料的使用以女性的資料為主，以育齡婦女的資料為主，通常不管男性的資料。雖然政府發佈的統計也有男性生育率的資料，但這在研究的作用並不太大，因此，是否要把男性的生育資料放進資料庫，值得考慮。

4、婚育資料庫的架構與內容：

話說回來，婚育資料庫中是不是得列入生命表？生命表的計算是以死亡率的估算為基礎的，而計算死亡率的精算式有 1000 多條，你列出的生命表是以哪一條的精算式為基礎？是統統列入？還是乾脆根本不要把生命表放進婚育資料庫之中？如果不列入生命表，那麼當然疾病的資料也不必放進婚育資料庫裡面。

關於婚育資料的具體項目的問題，台灣的婚姻資料並不完整，有偶率的資料缺 1976 年以前者，有偶人數的靜態資料缺 1964 年以前者，這些基礎的資料都是很重要的。

另外，同居生育的資料很重要。目前有偶婦女的生育率統計數據是千分之 1225。假定 15-19 歲的婦女每年皆生一個孩子，則未婚生育的比例還是有 1/4。但事實上，15-19 歲的婦女不會每年都生小孩，因此，未婚生育的比例應該更高。所以，內政部不能再把千分之 1225 的數據歸成千分之一千。同居應該變成正式的統計資料項目，而政府蒐集同居資料不應該有問題，不能只根據戶籍登記的婚姻數字來計算有偶婦女的生育。

三、林季平：

1、初級資料與次級資料：

初級資料與次級資料之區分並沒有明確的定義，建資料庫時，只要將資料的來源明確地註明即可，不必刻意區分。

2、資料庫維護與下載方式的問題：

一個重要的問題是，資料庫建成之後，誰來維護？一個資料庫要保持其壽命與功能，必須能不斷地更新資料。所以，在建資料庫及編寫有關的程式之時，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，或是讓使用者能夠從資料庫中直接得出其需要的各項變數的基本統計量。

近十五年，實務界和理論界的研究方向是：資料庫走向個體資料（micro data），傳統的總體資料（macro data）的作法不行，因為沒有彈性。現在運算的要求越來越精細，macro data 的結果常常做不出來，

或者很粗糙。為了精細的研究要求，你必須能自由地組合原始資料，所以你非需要原始資料不可，這也是美國自 1950 年代以來發展的趨勢。當然原始資料整合的問題就涉及的隱私權和安全的問題，這需要政府各部會的協調，並釐清法律權利的問題。

3、資料取得的法律權利問題：

目前的婚育資料，不論是公務統計還是調查統計，並不全由內政部產生，其他部會可能各自產生一些婚育資料。各部會有時有其本位主義，但這不見得完全不恰當，有時候是為了隱私權保護的問題，例如財政單位對財稅資料的保密。因此，規劃資料庫時，應考慮資料取得的法律問題，否則萬一事先規劃了很廣的資料來源，但日後碰上資料整合的法律障礙，資料庫的建立很難推動。

4、資料庫的內容：

現在的婚預料庫規劃案只是規劃，所以未來的方向很有彈性。陳信木教授提到「台閩地區人口統計」及其他人口統計資料的電子化，我認為這很重要的，這事也許不可能馬上做到，但規劃資料庫時仍應有此理想。

四、許汶瑛

1、資料取得的法律權利與初級、次級資料的問題：

林季平教授提到資料取得的法律權利的問題。目前凡是經政府各部門加總處理而產生並發佈的資料，在使用上都沒有法律的問題，有法律問題的是未經加總處理的原始資料，這裡面有個人的 ID，有個人的隱私，其隱私受到「個人資料保護法」的保護，例如，主計處向財政部、健保局取得資料時，其中的身份證字號都變成亂碼；主計處的

工商普查需要財政部的財稅資料，財政部也不直接提供，而要求主計處人員到財政部去作業。另外，主計處第四局把人力資源調查的資料提供第三局時，也都經過事先的加總整理，先跑出時間數列的資料，再交給第三局納入統計資料庫。因此，想要利用身份證號碼把各部會的原始資料結合起來，在法律和實務上來說是不可能的。婚育資料庫的建置仍應以業已經過統計加總處理的所謂次級資料為主。

五、陳信木：

1、生命表、死亡率和人口預測資料：

關於生命表和死亡的資料問題，如果主要仍是連結其他的生命表和死亡資料庫，而婚育資料庫僅提供簡單的壽命資料的查詢，則功能不大，不需要放進婚育資料庫。

至於人口預測的資料，也是類似的問題。經建會大約每三年就會預測一次，因此重要的不見得是現在這一版本的人口預測，而是每一次預測的結果。這就涉及到婚育資料究竟要放多少東西的問題。

依照現在婚育資料庫的規劃來看，架構十分完整。但是，如果把這些資料都納入的話，結果會像是國家的人口檔案。這便涉及到很重要的一個問題，預算的問題，要花多少錢？別的不談，即使只是人口統計的電子書，光光是過去各年「台閩人口統計」的電子化，我估計至少要花 4000 萬元。另外，大家未提到的，跟婚育有關的過去家庭計劃的資料，要不要納入？繼續擴大下去的話，要花多少錢？所以，我們要考慮建婚育資料庫的原來目的是什麼？最小範圍的婚育資料庫要花多少錢？如果目標太高，規劃案做出來，但最後沒有錢可做，則我怕這個規劃案最後也只能放進國家檔案裡面，無法實行。

六、莊義利：

1.初級資料與次級資料：

所謂的初級資料通常是一筆筆過錄出來的個人資料，不管是戶籍的、健保的還是調查的個人資料。根據這些個人資料的紀錄，我們可以算出一些指標，變成計算好的、整理好的次級資料。我認為婚育資料庫不必考慮那些原始的初級資料，因為使用原始資料必須經過一些處理的程序，必須經過計算以產生各種指標，這需要人力的成本，不利於婚育資料庫平常的維護。各部會有責任根據原始資料算出各種指標。婚育資料庫不必放進原始的資料。

現在有線上運算的方法，將查詢者的電腦直接連上原始資料，然後利用線上運算的程式，依據原始資料直接算出各種指標，而不必事先製作千百個 frequency 表、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。不過，這應該是各個政府機關內部的事。婚育資料庫是統合的資料庫，它所應該做的，就是根據各政府部門每月、每季或每半年更新後發佈的整理過的指標和結果，整合進婚育資料庫裡面，定期更新。

2、婚育資料庫的架構與內容：

關於婚育資料的大小和範圍、其原始目的為何，要從使用者的角度來看。首先，就資料的新舊來看，應先考慮最新的資料，歷史的資料要放多少，可以再討論。

陳信木教授提到，單單是台閩地區人口統計的電子化，就得花費龐大，但婚育資料庫的建立應從使用者的觀點來看，就婚育問題而言，使用者通常會用到哪些指標，這些指標和項目的選擇，應該是婚

育資料庫最重要的事。從台綜院規劃的架構來看，性別、年齡等是分層變項，而使用者對婚姻關心的通常是結婚率、離婚率、有偶率、對婚姻的態度和看法等等；對生育關心的是生育率、生育態度、理想子女數等等。這裡面有些資料（結婚率、有偶率）是內政部產生的加總資料，有些（例如婚姻態度）是衛生署或主計處調查的個體資料（micro data）。重點是，對使用者而言，婚姻資料庫應該有那些項目。如果把重點放在這裡，婚育資料庫也不會過於龐大。陳信木教授提到「家庭所」的家庭計劃資料，其實這些資料有些部份重要，有些部份不重要，真正篩選出來，項目也不是那麼多的。

七、楊文山：

1、婚育資料庫國際化：

此一計畫的目標之一，似乎包括與國際婚育資料庫（例如聯合國婚育資料庫、歐盟的婚育資料庫）的比較。因此，建立此一婚育資料庫時，應力求國際化。

2、研究功能與政策應用：

從目前的架構來看，婚育資料庫的規劃似乎偏向供研究者使用。但資料庫應有政策應用的功能，方便廣大的民間和政府部門使用者利用，應該做得更實用一點。

3、婚育研究資料庫：

規劃中的婚育資料庫主要是統計數據資料庫，但是，同樣重要的是婚姻資料的研究成果和婚育研究的文獻，這可讓人掌握婚育數據的政策意義。婚育研究成果的建置是另一種資料的內容，或許將來也可以納入婚育資料庫中。

4、資料庫的線上運算：

現在中研院調查中心的胡克威先生利用英國 Essex 大學提供的技術，建立個體資料（micro data）線上運算的方式，這種軟體提供變項的分配次數（frequency）計算和交叉分析等基本運算功能。使用者在使用的終端鍵入變項的名稱（例如有偶婦女生育率），便可得出此一變項的分配次數。婚育資料庫的建立或許也應往這一方向發展。

八、徐美：

1、現有婚育資料庫的不足：

現有婚育資料庫有時不能充分滿足研究的需要。例如主計處的「婦女婚育與就業調查」內有幼兒托育的問項，但僅有有偶婦女的調查，沒有離婚、喪偶者的托育情形，這使我們很難看到婦女婚育與就業之間關係的全貌。所以在規劃資料庫時，要考慮資料庫和分層變項的完整，並能隨著經濟的發展和變化而增加新的問項與資料。

2、婚育資料庫的大綱與細節

有時候看到一些婚育、人口的國內總體指標，再回去了解個體的資料（micro data），不免覺得國內的一些總體指標和個體資料之的關係不是很清楚。因此，建立婚育資料庫也許應該該先規劃一個骨幹、一個 map，然後先檢討大綱的設計，再填入細節，看一看資料庫需要那些具體的內容，那些是可以立即放進去的？那些是短期內部不見得做得到，但可當作長期的目標，也是內政部和主計處將來可以投入資源去蒐集建置的。

3、與其他資料庫的連結

對於有些不放進婚育資料庫的重要資料，應該以連結的方式處理，婚育資料庫應同時說明那些資料庫的使用者必須具備什麼身份，必須以何種程序取得那些資料庫的使用權利。

- 4、若要使婚育資料庫能夠做線上運算，可能需較多的經費，短期內大概做不到，也許可以在未來嚐試。以後的技術可能更成熟，運算速度更快，這可以是長期的目標。

九、陳子和：

在這裡對同居與婚姻資料的問題稍做說明。在戶籍登記的登記婚資料裡面，不包含同居的 couple。但在戶口普查時，同居被當作事實婚計入結婚的對數中。

十、楊靜利：

1、初級資料與次級資料的問題：

依照我們的規劃，婚育資料庫中放入的資料都是經過整理的統計結果，也就是次級資料，而不是原始的登錄形式或登錄碼。對於原始的初級資料，也許可以連結的方式，接到原始資料的所在（例如中央研究院調查研究專題中心），並在婚育資料庫中說明申請原始資料的程序。

2、死亡、生命表、人口預測的問題：

關於死亡資料、生命表該不該放入婚育資料庫，我想嬰兒死亡率、生產期間的婦女死亡率應該是可以放進去的，因為這和生育有關。

另外，陳敬宏統計長也提到，生命表和人口預測有關，所以還是可以放入婚育資料庫裡面。

3、調查統計資料：

現在政府部門的婚育統計資料有一些不足的項目，例如沒有按工作年數別區分的離婚率。這些缺失，或許可以利用調查統計（例 KAP）的資料來補充。陳敬宏統計長提到對調查統計資料之品質的審核問題，某些調查統計的資料，或許需要內政部另以專案的方式計算出結果，再經過審查，確認其品質之後，再列入婚育資料庫之中。

另外，婚育資料庫也許不必將調查統計單獨列出，只要將各類有關資料列在人口、婚姻、生育、社會經濟等各大項之下，再說明資料的來源即可。

4、資料庫的架構與內容：

各位對於婚育資料庫的涵蓋面、大小的取捨或有不同的意見，我想我們在婚育資料庫規劃的報告，可以提出分階段的建議：第一階段做什麼，第二階段做什麼等等。

十一、陳筆：

1、初級資料與次級資料的問題：

我們原先的構想是：婚育資料庫以經過加總整理的統計資料為主，暫時不做原始資料的整合。

2、資料庫維護與更新的問題：

我們同意林季平教授的看法，建立婚育資料庫要考慮後續的補充、更新和維護的問題，必須建立程式，使維護者能方便地更新。我們將對這一個問題提出建議。

拾、散會（上午 12 時）

