

《台灣社會變遷基本調查》 串連資料介紹

羅婉云

由中央研究院人文社會科學研究中心轄下調查研究專題中心所執行之《台灣社會變遷基本調查》跨年資料串連計畫，目前已釋出「家庭組」、「社會階層組」、「大眾傳播組」、「宗教組」等主題自二期至五期調查的串連資料，以下將針對串連目的、串連方法與釋出資料內容作一概述。

一、前言

台灣社會變遷基本調查（以下簡稱「社會變遷調查」）的同一題組或題項可能在不同期次的調查中重複出現。透過不同期次調查資料的串連，可得到重複橫斷面資料（repeated cross-sectional data）。對於研究者所關心的現象，這類資料不僅可用來進行跨時代或跨世代分析；也可應用某些統計方法解析出年齡、出生世代及時代的影響效果。對社會科學研究者而言，重複橫斷面調查資料可說是彌足珍貴的研究素材。

然而，相關的研究成果，都仰賴個別研究者自行串連自己所需的資料。個別研究者所串連的資料，固然可供做特定主題的分析素材；但每位研究者必須就自己的研究旨趣，投入相當心力在繁瑣的串檔工作上，造成研究資源不必要的重複消耗，及研究者額外的負擔。因此中央研究院人文社會科學研究中心調查研究專題中心建置的「學術調查研究資料庫」（Survey Research Data Archive, 以下簡稱 SRDA），自 2009 年起著手進行「社會變遷調查」的資料串檔工作。這些已串好的資料檔，目前放置在 SRDA 網站以及 Nesstar 線上分析平台，供研究者線上分析使用¹。儘管 SRDA 已有初步的串檔成果，但串連完成的檔案無論在主題或期次上，仍有侷限性。在前項工作著手進行後，也讓調查研究專題中心的同仁意識到，要進行「社會變遷調查」的跨年資料串檔，實在是非常艱鉅的工作。即使同一問卷主題（如家庭組、階層組）中的相同題項，在跨期次資料的串連上，已面臨多重挑戰。例如，同一主題在兩個不同期次調查中的題目，可能題意近似卻不完全相同，或題目相同但選項不同（如測量尺度不同、選項文字內容不同）。因此，即使以初步串好的三至五期家庭組、階層組、綜合組資料，要增進串連資料檔的可用性，仍必須從題項設計原意、使用價值等觀點，重新思考資料的串連方式。

有鑑於此，《台灣社會變遷基本調查》跨年資料串連計畫的目的是針對社會變遷調查最常出現的四項主題（家庭組、階層組、大眾傳播組、宗教組），分別進行跨期資料的串連。所串連的資料，包括主題輪替期次的題項，及各主題問卷中出現

¹網址為 <http://140.109.171.222/webview> 點選「中文」→「長期主題型調查」或「國際合作調查」→「台灣社會變遷基本調查」→「合併檔」

的個人基本變項資料。後者包含受訪者個人的人口特徵變項（如性別、年齡、宗教、出生地區、居住地區、教育程度、婚姻狀態、家庭規模、健康狀況、工作狀態、行業、職業、個人收入）、受訪者家庭的相關變項（如家庭收入）、受訪者父母的人口特徵變項（如父母親的籍貫、教育程度）、受訪者配偶相關變項（如籍貫、教育程度、工作狀態、職業）等。儘管這些變項在橫斷面調查資料的應用上，通常作為控制變項使用，但二十多年累積的資料若經整合，資料本身即是非常珍貴的研究素材，可用以探討社會人口特性的變化，及人口變項間的關係。此外，串連檔中仍保留各期次的原始變項資料，讓資料使用者保有資料使用的彈性，可自行將資料做進一步的整併、處理。

二、串連方法介紹

社會變遷調查是由行政院科技部長期支持的全臺抽樣調查計畫。自 1984 年開始進行調查，發展至 2012 年為止，已完成 23 次 48 份問卷的抽樣調查，調查內容包括家庭、教育、社會階層與社會流動、政治文化、選舉行為、傳播、文化價值、宗教等。社會變遷調查的樣本數相當龐大，每份問卷的完訪數約在 2,000 案左右。依照調查計畫的規劃，自第二期開始，有幾項大型主題是採輪替的方式，每隔五年進行一次調查。亦即，採用重複橫斷面調查的問卷設計方式。考量先前整併社會變遷資料的經驗，本次在規劃過程中，分別自資料庫管理者與資料使用者的角度，構思如何進行資料串檔；在實作中再逐步確認、修正題項整併的流程。以保留最多可用資訊，並讓資料的應用價值達到最大，為主要的處理原則。依照串連步驟，將成果說明如下：

（一）主題年題目彙整

歷年社會變遷調查，每份問卷的題目皆超過百題，要統整多份問卷題目，工程非常浩大。為了方便日後比對與排序，並有效找出相同（相似）的題目，先就四大主題（家庭、社會階層、大眾傳播、宗教）各調查期次（共五期 17 組）的所有問卷題目與選項著手²，將各主題問卷的題目及其選項分別彙整成四份 Excel 檔案格式的主題題目總表。在各個主題總表中，每個題目所包含之欄位資訊如下：

1. 編號：將資料檔中的題目（變項）依調查時間的順序編排流水號，藉以判別各期資料檔中的變項，並便於計算變項總數。
2. 主題組別：此欄位標示該題目所屬的調查主題為何，調查主題包括家庭組、社會階層組、大眾傳播組、宗教組。
3. 期次：題目所屬調查的調查期次。
4. 調查年度：題目所屬調查的調查年份。

² 三期三次社會階層組問卷分長卷與短卷。

5. 變項名稱：在原始資料檔中的變項名稱。
6. 變項說明：變項的原始問卷題目，可作為界定題目分類第一碼與第二碼欄位（參見後文第（8）至（10）項）之依據。
7. 選項說明：原始問卷題目的選項，列出各選項中的編碼及其代表意義。
8. 題目分類：根據問卷題目在原始問卷中標示的子標題，作為此欄位分類的參考。例如，「請問您目前信什麼教」這題，在原始問卷中的子標題為「宗教信仰」，即是此欄位的內容；「現在或退休前，您幫公家還是私人工作」題目，在原始問卷中的子標題為「職業狀況」，即為此欄位內容。其餘依此類推。
9. 題目分類碼一：此欄位是將相近的題目大類給予同一編碼，方便在題目排序整理時，可以將相近題目放置於同一編碼下比對。例如，歷次的問卷中，與教育狀況相關的題目皆給予代碼 02，與婚姻狀況相關的題目給予代碼 03，與宗教信仰相關的題目給予代碼 04，與居住環境與條件相關的題目給予代碼 05，與職業狀況相關的題目給予代碼 06。
10. 題目分類碼二：此欄位內容是將前項（題目分類碼一）相同的題目進行跨期次整合，依題目相同與否歸類排序，再以流水號的方式逐一編碼。舉例來說，題目分類碼一「02、教育狀況」中，第一題「請問您的教育程度是」，本題於階層組二期三次至五期三次的四個調查中出現，且題目相同，因此這四個變項在分類碼二的欄位皆賦予相同的編碼（001）；第二題「請問您是否有完成學業」，出現在階層組四期三次至五期三次的兩個調查中，則此兩個變項在分類碼二的欄位皆賦予相同編碼（002）；以此類推。

最初進行主題總表規劃時，只針對題目進行第一碼的分類，但經過排序比對後，發現只有一碼的分類仍不足以標示题目的相似程度，因此再進行細類的第二碼編碼。過程中經過多次的測試與討論，以求分類方式能易於操作且能提供足夠的資訊。前述第一碼與第二碼的編碼原則，便於工作人員分辨哪些變項是相同的，以將編碼相同的變項整併為同一變項。

另外，在分類過程中，發現不少題目近似，但題意或問法不盡相同的情況。處理方式是將其視為不同題目，於分類碼二的欄位中新增一碼，如表一所示。對這類題目相近但不完全相同的變項，並沒有整併為同一變項，主要的原因是考量到使用者可能有不同的研究目的。如果逕自整併為同一變項，可能與研究者的構想不符，反而讓研究者在使用串連資料時，必須花額外的時間再做處理。

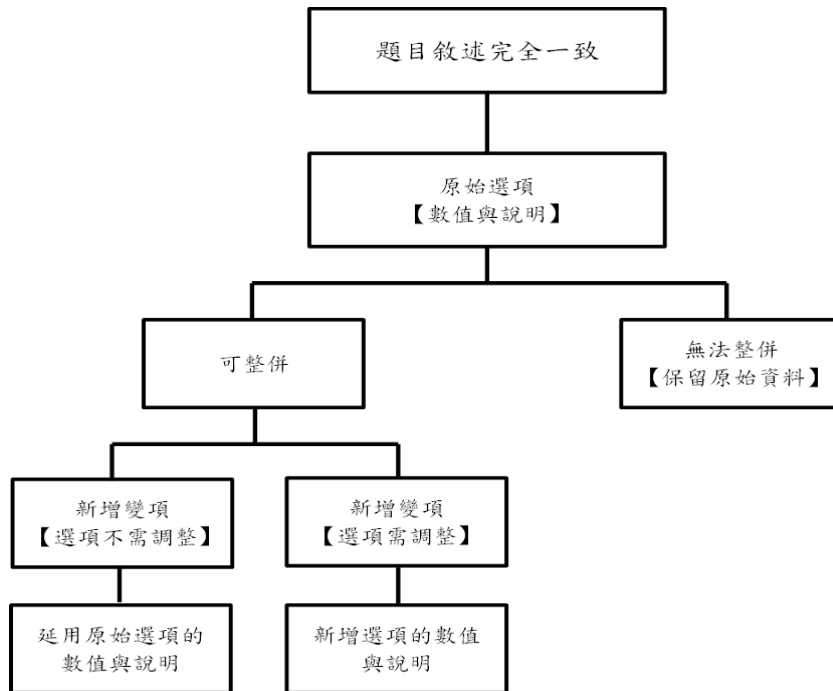
如上所述，SRDA 只對題目完全一致的題項加以整併，對於題意近似但不完全一致的題目，並未予以整合。但串連檔中保留了原始的題項資料，可供研究者自行決定處理方式。

表一 主題總表：題目「分類碼二」編碼範例

編號	主題	期次	調查年度	變項名稱	變項說明	選項說明	題目分類	分類碼一	分類碼二	整合
113	宗教	0305	1999	v22	您認為自己是不是 一個還算 虔誠的宗 教信徒？	(1)是 (2)不是 (7)不知道	宗教 信仰	08	052	否
257	宗教	0405	2004	v22	請問您認 為自己是 不是一個 虔誠的宗 教信徒？	(1)是 (2)不是 (7)不知道	宗教 信仰	08	053	否

(二) 選項整併

四個主題總表完成題項分類後，根據各表中「題目分類碼二」欄位資訊，選擇分類碼編號一致的題項，從這些題項中開始進行選項細部的比對工作。選項的整併方式，分為三種情況。首先，若題項的選項一致，則確認此為可整合的變項，並於表中的「是否整合」欄位標記。其次，若題項的選項不一致，但仍可歸類並重新編碼為相同類別，則視為可進行選項整併的變項；於主題總表的備註欄中註記相關資訊，提供檔案串連時撰寫變項合併程式的參考。最後，若題項相同但選項定義不同或測量尺度不同，導致選項無法整併，則不予以整合；但仍於串連後的資料檔中保留原始資料，讓使用者在分析資料時，自行決定如何運用。選項整併的基本架構，如圖一所示。



圖一、選項整併架構

(三) 檔案串連

根據主題總表所整理出的資訊，開始進行四大主題調查資料的串連工作。如前所述，為了讓使用者在分析資料時有較大的彈性，在進行資料串檔時會於資料檔中保留所有的原始變項。為了達到這項目標，且為識別串連檔中變項的調查期次來源，在各主題的資料進行串檔之前，先將所有變項重新命名，於變項名稱中加入調查期次資訊。以階層組為例，二期三次至五期三次共五份問卷資料，在進行資料串檔前，先於原始變項名稱前增加若干字元，包含字母 q、調查年份（西元）末兩碼、問卷別（問卷一、二分別以數字 1、2 代表）、長短卷別（長、短卷分別以英文字母 l、s 代表）、下底線，如表二所示。重新命名的目的，是讓使用者能從串連檔中的變項名稱，一目瞭然其所屬的調查期次等資訊，不須耗費額外時間查詢。

表二 串連資料檔變項名稱命名範例

階層組問卷期次	調查年份	原變項名稱	串連檔變項名稱
二期三次問卷一	1992	v1	q921_v1
三期三次問卷一長卷	1997	v1	q971l_v1
三期三次問卷一短卷	1997	v1	q971s_v1
四期三次問卷一	2002	a1	q021_a1

階層組問卷期次	調查年份	原變項名稱	串連檔變項名稱
五期三次問卷一	2007	a1	q071_a1

四個主題資料經串連後，完成的家庭組串連總檔共計有 1,892 個變項，樣本數為 8,493；階層組串連總檔共計有 2,711 個變項，樣本數為 10,722；大眾傳播組串連總檔共計有 1,837 個變項，樣本數為 8,006；宗教組串連總檔案共計有 1,706 個變項，樣本數為 7,595。由於社會變遷基本調查釋出的資料內容可能隨時間而更新，茲將串連時所使用的各期次原始資料檔資訊（含資料釋出時間、變項數、樣本數等）列於表三。

表三 串連檔資料來源說明

調查主題	問卷期次	調查年度	資料檔版本 (釋出時間)	變項數	樣本數
家庭組	二期二次問卷一	1991	2013.04	352	2,488
	三期二次問卷一	1996	2012.10	541	1,924
	四期二次問卷一	2001	2010.09	570	1,979
	五期二次問卷一	2006	2013.07	429	2,102
階層組	二期三次問卷一	1992	2013.03	572	2,377
	三期三次問卷一長卷	1997	2012.07	495	2,596
	三期三次問卷一短卷	1997	2012.07	257	1,717
	四期三次問卷一	2002	2010.09	631	1,992
	五期三次問卷一	2007	2013.07	756	2,040
大眾傳播組	二期四次問卷一	1993	2013.01	362	1,946
	三期四次問卷一	1998	2011.02	395	1,919
	四期四次問卷一	2003	2010.06	531	2,161
	五期四次問卷一	2008	2013.10	549	1,980
宗教組	二期五次問卷二	1994	2013.01	547	1,862
	三期五次問卷二	1999	2012.07	443	1,925
	四期五次問卷二	2004	2012.07	326	1,881
	五期五次問卷二	2009	2013.05	390	1,927

(四) 變項整合

依據先前四大主題總表所整理出的題項細部資訊，進一步進行四個主題資料檔的跨期次變項整合工作。由於保留了原始變項的資訊，每個串連總檔皆包含上千個原始變項。其中，階層組由於三期三次調查同時有長卷與短卷資料，串檔後的總檔

變項數更高達 2,711 個，需要處理的變項數目眾多。另外，在跨期次資料的串連過程中，也必須考量到題項與選項進行合併的合理性與複雜度。故在撰寫變項合併程式時，比對程序十分繁瑣。由於先前已整理出四個主題總表資訊，使得實際撰寫變項合併程式時能事半功倍，減少在整合過程中發生錯誤的機會。對變項整合過程中的處理方式，列點說明如下：

1. 新增變項 (variable)：串連檔除保留各期次原有變項外，另將總表中註記可整合的題項新增為另一變項，利用合併程式予以整合，並將新增的變項名稱重新命名。如同主題總表的整理原則，新增的整合變項是針對出現於兩份或兩份以上問卷（可能為不同期次，或是同一期次的長短卷）的相同題目，才加以整合。如果題目僅出現在單一問卷，或是題意相近但題目的文字並不完全一致，則沒有整合的問題，但仍會將這些題項原始資料保留於串連檔中，以求主題檔案資料的完整性。
2. 整併選項 (value label)：對於新增整合變項的選項整併，依情況不同有三種處理方式。首先，若變項在各期資料中的選項一致，在進行變項合併時，新增整合變項的選項編碼，會直接套用最近一次調查資料中的選項編碼。其次，若新增整合變項之選項內容，在不同期次的資料中僅略有差異（如選項編碼不一致，但文字意義相同），則視為可整併的變項，仍可利用重新編號 (recode) 的程式語法加以整合，並套用最近一次調查資料的選項編碼，作為新增選項之編碼依據，如表四所示。

表四 選項整併範例（一）

問卷期次	原始選項	整合後之選項
二期三次問卷一	1 很同意 2 同意 3 無意見 4 不同意 5 很不同意	1 很同意 2 同意 3 不同意 4 很不同意 5 無意見
四期三次問卷一	1 很同意 2 同意 3 不同意 4 很不同意 5 無意見	

最後，若變項在不同期次資料中的選項類別或尺度不同，即使題目相同，亦不予整合。舉例來說，宗教組的題目—「最近一年內您有沒有安太歲」，雖然三期五次與四期五次調查皆有此題目，但在選項類別與意義上略有不同，故不進行變項合併，如表五所示。

表五 選項整併範例（二）

問卷期次	變項名稱	變項說明	選項說明	整合
三期五次 卷二	v67c	最近一年內您有 沒有安太歲？	(1)有 (2)無	否
四期五次 卷二	v46c	最近一年內您有 沒有安太歲？	(1)自己 有做 (2)自己 沒有 (3)自己 沒有， 家人有	否

3. 新增變項之命名原則：對新增變項，為方便研究者在使用串連檔時能容易瞭解變項意義，在新增之整合變項重新命名時，特將變項名稱進行有系統的命名與排序。依照變項屬性分為兩種情況，其一，若屬於人口特徵變項，則以題項意涵命名。例如，性別的變項名稱為 sex，教育程度的變項名稱為 edu，婚姻狀況的變項名稱為 mari ... 等等。若為工作狀況題組，變項名稱以 work 起始，再加上 3 位數的流水號以區別變項，例如 work001、work002、work003 ... 等。若為休閒生活題組，變項名稱以 lei 為首，再接續 3 位數流水號，例如 lei001、lei002、lei003 ... 等。四個主題串連檔的人口特徵變項名稱，如表六所示。

表六 主題串連檔人口特徵變項名稱對照表（節錄）

編號	變項名稱	變項說明	有串連變項			
			家庭	階層	傳播	宗教
調查資訊						
1	data	期次卷	√	√	√	√
2	id	樣本編號	√	√	√	√
3	year	調查年(西元)	√	√	√	√
4	year_m	調查年(民國)	√	√	√	√
5	wave	期次別	√	√	√	√
6	qtype	問卷別	√	√	√	√
7	zip	地區郵遞區號	√	√	√	√
8	weight	權數	√	√	√	√
個人基本資料						
9	sex	性別	√	√	√	√
10	biry	出生年(民國)	√	√	√	√
11	birmonth	出生月	√	√		√
12	edu	受訪者教育程度	√	√	√	√

編號	變項名稱	變項說明	有串連變項			
			家庭	階層	傳播	宗教
13	graduate	有無完成學業		V		
14	fedu	受訪者父親教育程度	V	V	V	V
15	medu	受訪者母親教育程度	V	V	V	V
16	sedu	受訪者配偶教育程度	V	V	V	
工作狀況						
67	work001	您現在在那工作?為誰工作	V	V	V	
68	work002	您是否與人合夥?	V		V	
69	work003	您有沒有雇人?	V	V	V	
70	work004	您現在雇多少人?	V	V	V	
71	work005	您現在的工作雇用的這些人和您有何關係?	V	V	V	
72	work006	您幫公家還是私人工作?	V	V	V	

若為四大主題相關的變項，則依照該主題之屬性命名。家庭相關題組，變項名稱以 fam 開頭，再加上 3 位數流水號，例如 fam001、fam002、fam003 ... 等。階層相關題組，變項名稱以 stra 開頭，再加上 3 位數流水號，例如 stra001、stra002、stra003 ... 等。若為大眾傳播相關題組，變項名稱以 com 起首，再以 3 位數流水號依序命名，例如 com001、com002、com003 ... 等。若為宗教相關題組，變項名稱以 rel 起首，再以 3 位數流水號依序命名，例如 rel001、rel002、rel003 ... 等。

4. 統一選項的特殊碼：由於社會變遷基本調查資料囊括諸多調查主題，且調查年份橫跨多年，歷次資料在特殊碼的設定上，難免出現不一致的情況。故將四個主題串連檔中常見的選項特殊碼予以統一，並盡量維持兩碼的設定，如 96 代表「漏答」、97 代表「不知道、不記得」、98 代表「拒答」、99 代表「不適用或跳答」等。
5. 不合理值檢核：串連各期次檔案的過程中，同時也會進行資料不合理值的檢核。如果發現原始資料檔中的變項說明（variable label）或選項說明（value label）有錯誤或缺漏的情況時，即依據該期次的問卷資料內容，補充該錯漏項目，如下頁表七所示。

表七 串連檔變項說明勘誤範例

問卷期次	變項名稱	變項說明 (variable label)	
		原始	更正後
三期五次卷二	v66f	您相信個人命運可透過重修組墳改變嗎？	您相信個人命運可透過重修祖墳改變嗎？
四期五次卷二	v81	請問您認為中國悠久的歷史值得驕傲嗎？	請問您認為中國悠久的文化值得驕傲嗎？

由於社會變遷調查計畫在每次調查完成後均會進行邏輯檢核，對串連檔中題目間的跳續答邏輯進行一致性檢核，並未列入本次串連之工作範疇。在變項合併的處理上，以保留原始資料檔之選項答案為原則。舉例來說，大眾傳播組題目—「去年一年內您平均為自己買了幾本書」，在二期四次至四期四次調查資料中皆有此題，惟不同期次的問卷題目不盡相同。譬如四期四次問卷，在此題之前尚有題目：「去年一年內，您大約多久看一次書」，對該題選取「(08)去年一年都沒有」選項的受訪者，接下來的「去年一年內您平均為自己買了幾本書」這一題會被設定跳題，其選項答案設為「(99)跳答或不適用」。但在其他期次的問卷中，並沒有這項跳題設定，對回答「去年一年都沒有看書」的受訪者，資料檔中的答案會是「(0)0本」的選項。

跳題設定會因各期次問卷設計而有所差異，考量到各期資料在跳答設定上的複雜性，因此在合併各期變項時，除非是整併選項的需要，否則會盡量維持原始資料中的選項數值及其說明。以前述題項為例，並不會將四期四次資料檔中的「(99)跳答或不適用」選項答案改為「(0)0本」。若使用者因研究需要而想將變項資料重新編碼，可參酌釋出檔案中所附之各期次問卷檔案，自行處理資料。

經整合後，四個主題串連檔中的新增變項數，家庭組新增 124 個變項，階層組新增 408 個變項，大眾傳播組新增 223 個變項，宗教組新增 346 個變項。

三、釋出文件

串連完成的資料檔，將放置在 SRDA 網站與 Nesstar 線上分析平台，供研究者下載或線上分析使用，欲瀏覽串連資料的釋出項目，請至 SRDA 網站的「長期主題型調查」中點選「社會變遷加值資料」進行瀏覽，或登入會員後下載取得相關檔案。

雖然社會變遷調查在釋出歷年資料時，並未釋出過錄編碼簿，考量串連檔案的複雜性，SRDA 另行製作四個主題串連資料檔對應過錄編碼簿，於釋出檔案中提供使用者下載。釋出資料除了過錄編碼簿與串連資料檔外，並附上串連檔所使用的各期問卷檔案、串連程式、與串連變項對照表、串連變項次數分配表等相關文件，供

使用者參考。以階層組主題為例，釋出資料內容如表八所示。

表八 階層組串連資料釋出檔案內容

釋出檔案名稱	內容說明
1. 各期次問卷檔	
tscs921.pdf	二期三次問卷一
tscs9711.pdf	三期三次問卷一長卷
tscs971s.pdf	三期三次問卷一短卷
tscs021.pdf	四期三次問卷一
tscs071.pdf	五期三次問卷一
2. 串連資料檔	
stratification_v01.sav	SPSS 資料檔
stratification_v01.dta	STATA 資料檔
3. 串連程式檔	
stratification.sps	跨期次重複題項串連程式檔 (SPSS v20.0)
4. 串連資料過錄編碼簿	
codebook.pdf	串連資料檔之過錄編碼簿
5. 串連變項對照表	
階層組串連變項對照表.pdf	新增串連變項、原有變項之變項名稱對照表
6. 串連變項次數分配表	
階層組串連變項次數分配表.pdf	新增串連變項之次數分配表

完成的社會變遷調查串連資料，可大幅提昇社會變遷調查資料的價值，使得早期的調查資料，可因此而更加活化；而隨著未來社會變遷調查的逐年推展，亦可進一步將新的調查資料整合到串連資料中，進一步增加資料的附加價值。所建構的串連資料檔，不僅節省研究者自行串檔所須消耗的人力、資源，也拓展了可行的研究議題。透過這項串連計畫的成果所帶來的附加價值，不論調查資料本身抑或學術應用價值，都是相當可觀的。