

## Chapter 3:獲取空間數據

### 3.1 引言

要把空間數據輸入到 GIS 中，並沒有便宜和輕鬆的方式。在許多案例中，只要是建立 GIS 專案中的數據收集階段，是所有階段中耗時最長以及最昂貴的。BernHardsen (1999)提出，一個能完全操作性的 GIS，其收集與維護資料的花費，無論在金錢與時間，均佔總成本的 60~80%。其他作者則提出更高的數據，例如 Siebert (2000)在他建立的東京歷史 GIS 中，即表達出一份所耗時間與所接觸問題的真實花費。

空間數據有兩個基本來源：主要來源即資料直接輸入於 GIS 中。例如：使用全球定位系統，或是遙測衛星。對歷史學者而言更普遍的是：次要來源擷取，即資料由紙質地圖轉化成數據型式，有兩種方式可以達成此項：掃描地圖產生網格式數據；或者數位化地圖，即直接由紙質地圖或是直接由地圖影像，勾勒出點或線，此會產生向量數據。

值得注意的是，GIS 對主要與次要資料來源的定義，與歷史學家所用的定義不同。從 GIS 觀點，主要資料來源是由真實世界直接擷取，而次要資料是由抽象的來源擷取例如地圖。一份於 1850 年製作的地圖，由歷史學者的角度是主要來源，但由 GIS 觀點卻是次要來源。

另一個由自己來擷取資料的選擇，是由其他人處取得數據資料。

這是耗費較少時間，也較沒風險，但因空間數據通常很昂貴，可能基於他人的使用，會有嚴重的版權限制，而也需被修正以符合你所需要的目的。

鑑於歷史的研究，許多必須的基本數據，因缺少要求/需求，而沒有數字化。因此，通常無選擇的僅能由自己去擷取。

首先，也可能是最重要的是，當擷取數位空間數據時，需決定要用何種來源。這牽涉到數據是在屋內擷取，或由其他方式獲得。原始來源的限制，會連帶影響而限制到任何此數據要接續使用的。來源的規模，在此特別重要。一般而言，更大規模的資料，會比更小規模的更具彈性。但無論在購買價格或取得所耗時間也會更昂貴，並且也會有更多多餘的訊息而導致檔案大小的問題。第二個需考慮的事項是，原始地圖的製作，與任何代表性數據之目標，是否相容。

### 3.2 掃描地圖產生的數據資料

掃描是一個比較直接性的方式，只要將紙質地圖放置於掃描器上，然後製造出格柵式拷貝。小型的掃描器通常較不昂貴。然而適合大型地圖紙之較大型掃描器仍然很昂貴。掃描器的空間解析通常用每平方英寸數點表達，而其光譜解析(能分辨之色彩數量)必須緊記，此將會影響最終資料的準確性。從最終性的掃描轉換到第二章所述之網格式數據，將需要相當數量的後製處理。

假使來源地圖相對的簡單且專一於一個主題，例如土壤型態的土地利用，也許會相當的直接。而從較複雜的來源轉換會是非常費時的。

### 3.3 數位化地圖來製造向量資料

當掃描製出來源地圖之拷貝時，由來源數位化摘取適當的特徵且由其製成點線與或多層面。從早期的 Ordnance Survey(OS)英吋至哩地圖系列使用者，可能僅要將一層的點代表教堂的地點，其他的圖層如公路與鐵路網，行政區邊界，等等也能被摘取。然而，卻是不可能如掃描所做的，去由來源地圖製作直接的拷貝。

點是藉由點擊所需要的目標而數位化。線段則是藉由沿續著線並且在線明顯改變方向處點擊。平面則藉由在線數據上建立拓樸學結構而產生，如第二章所述。無論是使用數化桌或數化板，直接由紙質地圖，或先掃描地圖且在螢幕上數位化，通稱抬頭式數位化，均能完成數位化。數化桌與板

(板通常較大)通稱為數字化，由被緊貼的地圖上之平面所組成。板或桌之表面有一組精緻銅線網在其下，也有一個圓盤，一種擁有精緻交叉髮線結構及一個或多個按鈕之手握式設備。要摘取一個點，就將交叉髮線結構放置於感興趣的目標上，再按下按鈕。桌或板能計算交叉髮線結構的正確地點，而將相對資料傳予 GIS 軟體。抬頭式數位化並不需要數化者，就如同掃描已經數位型式的地圖。而是使用電腦滑鼠將遊標放置於螢幕上之目標，再點擊按鈕，以決定正確地點。抬頭式數位化具有能掃描來源拷貝，且能從中擷取向量資料之優勢。此既能保留來源之拷貝，且能被當成提供擷取向量特徵之環境背景所使用。

數位化準確性是非常重要的，尤其是當拓樸學被製作時。一個節點通常需要兩到多次的數字化，以代表一線段的終點以及另一線段的起點。大部份的 GIS 軟體對假使在一組距離內的兩節點，將會給予容限，而結合形成一個節點。假如容限設定太高，則不適當的目標會被結合。如果容限設定太低，則會出現空白，多邊形將不能適當結合，通稱為晃點。此將導致混淆的拓樸學

### 3.4 地理坐標參照

理論上，不論是否被掃描或數字化，這一階段的數據採集過程通常是以英吋或厘米實測獲得的基礎座標圖。那是從左下角的掃描器或數化器而得。地理座標參照，是這些座標經由投影系統轉換成真實世界座標的過程。這使得距離和區域的計算和數據是不同來源的整合。

地圖投影是錯綜複雜的。像英國是一個使用國家網格十分普遍的國家。對投影法的詳細解釋，認為不怎麼必要的。在這個簡短繪圖指南下，經由參考書目列出的工作，可以進一步發現細節。地球是球形物，目前世界上的點是，使用緯度(赤道南北向的數目經絡)和緯度(格林威治子午線的東西兩方之數目經絡)來描述。地圖是平面紙，投影是將彎曲的地球，轉換成平面地圖投影。此會引起一個或更多的距離，角度，區域或者形狀的變化。投影通常也因為轉換經緯度，角度，座標的經度和座標的緯度，特殊的點改變了英哩或公里。

英國的國家網格是一種麥卡托投影。其橢圓形球體中心在偏格林威治西方 2 度之南北向，恰位於英國中心脊髓。由這條線向西或東移動，距離會特別的增加變形。幸好，英國是一個長而薄，近乎南北向的國家。這些變型對歷史學家的操作較不具有顯著的規模。習慣上，英國被細分成邊長各 100 公里之正方形，每個正方形給予兩個字母的辨識標誌。而地區則由正方形西南角的公里或公尺距離所表示。例如 NN 是於蘇格蘭南方高地的正方形。對電腦來說，用字母編碼來使用是很笨拙的。而一個錯誤的起點被整個國家所給予。此點即在 Scilly 島西南方，使英國本島的所有地點，以不超過六位數之非負數整數的最近距離所表達。地點 253000，720000 是錯誤起點的東方 253 公里，與北方 720 公里，是在南方高地。這種使用六位數國家網格參考結構，對計算兩點之間距離變得很容易。

大多數的 GIS 套裝軟體使得地理索引過程看來較為直接利用來源地圖使用者發現真實世界相對於一個數量的參考點通常是 4 這些又被稱作 tic 點，通常為地圖紙的角落，這些點於是被數位化使用者則在將真實世界的座標輸入成點資料，軟體則利用這四個點轉換每一個座標變成圖層，通常軟體在這時也會產生投影系統，並同時轉換層面，這代表所有在層面上的點區會被以國家網格或其他被選擇的系統所合成，而所有在層圖上的距離測量將會以公尺公里或其他使用之參考單位表示。

使用有顯示座標網格且具現成可使用的投影資訊之現代地圖是較容易使用，但要使用不具座標網格的較舊地圖則可能會困難。此點能由原始地圖且在現代地圖也有設法找到參考點(適當的特點也許包括教會、燈塔、制動點或火車站)。然後這些點座標可能被發現，使用一個具座標網格的現代地圖。為求準確起見，現代地圖應該使用比來源地圖更大的比例，如果在地圖投影沒有資訊則譬

如 Delano Delano-Smith 和 Kain (1999), Harley (1975), Oliver (1993), 歐文和 Pilbeam (1992)的書是好的資料來源。對當地研究領域的工作, 投射的衝擊可能很小, 他們也許不是相關的。

### 3.5 誤差和精確度

正如前面所述, 對原始的地圖來說, 一份數位表示的紙質地圖是與原始地圖相同質量的好, 但它幾乎不可避免地將累積一些其他的誤差或不精確性。重要的是能區別不同的類型錯誤和不精確性。Unwin (1995)使用一個六個單向分類如下: 誤差用來區別在現實和數字表示法之間; 失誤則是根本的錯誤, 而準確性是精密的測量、結果、計算或估計、對價值的接受猶如真實一般; 精確度通常是在小數位的數量來更正可能證實的測量; 特性的目的是適合資料, 當使用資料時, 不確定性測量疑義或懷疑的程度。

尤其是掃描、數字化和地理參考資料消息來源所產生的位置誤差。數字化地圖也必須先完全地被安置平展在數字化儀或掃描器。即使這並不總是像聽說的那麼簡單: 地圖被摺疊存放, 也許隨時間而偏差變形等等。精確的掃描或數字化設備的準確性可能也是下個誤差的來源, 然而如果一般專業(通常是昂貴的)硬體只有輕微的小誤差, 假使由掃描器導致的錯誤而使用抬頭式數化器所漸增的錯誤, 最後數位化也會錯誤。

下一個誤差源是來自於用戶取得數據資料的過程。當人們獲取數位化的資料時, 必須在想要擷取的某一資料點上將游標移置該處, 即便是一個有高度積極性和機敏的人也會犯小小的定位誤差。當廉價的員工花上許多小時將資訊數字化, 不精確和失誤也會增多, 往往不能夠準確地擷取最確切的那點位置。點標誌並不總是精確地被定義為一點: 例如在鐵路車站, OS1: 50000地圖當中, 火車站幾乎是由2mm直徑的圈來代表。路線的圖徵加困難, 數字化線段有賴於操作員擷取線形變化方向的各點。至於平緩曲線, 譬如那些道路、河流或等高線, 這是一個無法避免的主觀性選擇, 因為不會有兩個操作員用此法數位化一條線時, 去使用相同點數位化去描述。

最後一個誤差源是地理資料參照的過程/進程, 從地圖所計量的座標將有一定量的誤差, 該地點的參考點位置也會有某些錯誤。這意味著, 在圖層上所放置的每個點都會有些微地扭曲變形。多數軟體套件, 將一定程度的誤差表示為「根均方(RMS)的誤差」, 這往往也體現在數字化儀單位和真實世界的單位上。ESRI 公司建議 RMS 誤差不應超過 0.003 數碼英寸 (ESRI 公司 1994a), 即使這個標準並不總是用在歷史的地圖上, 但重要的是, 在作為部分數據資料獲取的過程中, 它至少建立並且提供最低標準的準確性(參見第 9 章)。針對這個部分的關鍵是, 有許多相關錯誤和不確定性問題的空間資料, 歷史學家可能是不熟悉的。這些所引起的問題也許似乎嚇人; 然而, 由面對上述問題, 以及意識到他們蘊含的意義、應該不致於在研究過程中導致多麼嚴重的困境。

### 3.6 數字化屬性數據

這過程通常是比獲取空間數據來得更簡單, 屬性資料通常是經由掃描和光學字符識別(OCR), 或經由打字鍵入而取得。在 GIS 軟件裡, 這些屬性資料有的可以直接對空間特徵資料直接地鍵入, 或可分別輸入, 並且融入了空間資料以結合兩者的關係。對於更多關於如何獲得屬性資料的最佳操作資訊可參考 Townsend 等(1999)。

### 3.7 網格式對向量式以及向量式對網格式的數據轉換

對於獲取空間性數據資料來源的正義合法性問題 (Holy Grail)(註 1), 只要能夠掃描一張地圖, 然後從掃描數據自動地提取點、線、多邊形的特點, 這正是所謂以網格式至向量之轉換, 但並沒有證明很容易執行的。一些系統聲稱能夠做到這一點, 但過程中, 通常需要大量的使用者進行交流。即使如此, 它執行的結果仍可能是不準確的, 例如: 線框可有順序地影響接下來選取基本網格式格局

型式的操作步驟，且向量對網格的轉換可取得高度的成功機率，但還是需要注意一些採取措施。許多軟件包包括路由器，雖然會促使這兩項資料型式在該轉換的地方轉換，但也有許多對執行之潛在準確誇大不實。在一項重大購案或重大投資決定之前的重要時機之前，資料轉換的便易性與否是第一優先可信度價值極高的檢核條件，而不是只相信提供內部銷售攤位的軟件開發公司。

### 3.8 原始數據的來源

兩個主要的原始數據資料來源是從衛星圖像和全球定位系統（GPS）獲取而來的。衛星圖像是一種網格形式的數據，其中每個像素（pixel）代表地球表面的一部分。像素的確切尺寸取決於使用的掃描器和後處理的應用，這兩種資料處理過程的規格和功能，通常是一個兩側從 1 米左右到 100 米之內的方陣。每個像素數據也提供對於當地球表面某部分被衛星影像擷取後反射回來的光譜資訊。

再者，確切的細節在很大程度上依賴於掃描儀的類型和後處理的應用。從這些基本資料中，我們可以更进一步地建構、發展有關的知識，例如各類土地利用的技術、分析何處有健康的植被等等。不過在這裡並不打算對遙感衛星影像技術多做描述，因為在本章最後有收錄列舉多篇相當好的關於該領域的指南入門文獻，可供有興趣者進一步深入閱讀。

顯然，衛星影像在近年來才相對地能夠被普遍使用，但它卻是能提供給歷史學家相當有用的訊息，舉例來說，我們可以在某些遙遠地區中標立定義一些具特徵性的位置，或者試圖定位過去時空利用的土地範疇。從空中拍攝的照片中，我們可以瞭解在過去時空中土地利用的情形，並取得一個具選擇性的（二手的）重要統計資料來源。

全球定位系統接收器，是一種原始空間數據擷取的簡單方法。最簡單的使用形式就是全球定位系統接收器會給予使用者當下正確的所在位置並指引方向。這可以經由美國軍方所策置的衛星網路系統計算而得之。最初全球定位系統提供非軍方用戶使用的這些定位點準確性被蓄意地降低，誤差較軍方使用者差了有 100 米之譜。而這種情形隨著有其他國家投入發展太空衛星科技，和民間業界的 GPS 系統研發競爭下，如此大之誤差如今已經不復見，現在提供給所有用戶使用的 GPS 定位誤差值距離精確值只有數米而已。另外因為有更多精密的定位系統推陳出新，使用戶可經由 GPS 接收器獲取很多在天空中飛行的衛星訊號點，且接收器允許用戶下載訊號點到自己的電腦內直接存入。如果有需要更精準的準確度就要使用差動性（differentail）全球定位系統（DGPS）。該技術需要使用兩個接收器，其中之一接收器須保持固定在一個已知的位置，以協助另一個環顧巡迴在所處地點上空的接收器測量該地點的位置，而這仍有可能產生約分米的誤差值。

全球定位系統可以用來提供一個比較快速的方法進行位置測量和擷取地球地表上具有特別意義性的定點。有一個很好的例子就是 Lowe's (2002) 用衛星科技定位測量由戰爭遺址殘存下來的土木工程和其他具特別意義性的區域點，該區域點位於一座坐落於美國歷史上著名之南北戰爭當年戰區內的 spotsylvania 國家軍事紀念公園中。

### 3.9 收購或免費獲取數據

在英國學院，有各種各樣的社會團體提供許多種免費的或是低價的 GIS 資料。對歷史學者而言，最有用的是 AHDS History，<<http://www.ahds.ac.uk/history/index.htm>>。藝術與人文學科資料，<<http://www.ahds.ac.uk/>>。在愛丁堡大學，EDINA 提供包括 UKBORDERS，這是關於英國當代和歷史的資料，以及數位地圖。在曼徹斯特大學，MIMAS 提供多樣的資料，包括 Bartholomew's 地圖資料和衛星影像。英國國家檔案館（前身為公共記錄辦公室）：<<http://www.nationalarchive.gov.uk/>>，逐漸轉移到電子出版，目前可在網

上查到的只有數量有限的數據。其他免費或便宜的數據資料包括 Geography Network : <<http://www.geographynetwork.com/>>。還有 Digital Chart of the World : <<http://www.maproom.psu.edu/dcw/>>。提供各個國家不同的資料。還有美國地質調查 USGS，提供便宜的產品。<<http://www.usgs.gov/>>。The Electronic Cultural Atlas Initiative (ECAI) <<http://www.ecai.org/>>，e 化歷史資料並上線，也提供資料的 metadata。

有許多商業公司會出售的空間參照的數據。最明顯的例子 Ordnance Survey，這公司銷售許多數位化的地圖產品。其他商業資源包括機管局 AA <[http://www.theaa.com/aboutaa/data\\_sales.html](http://www.theaa.com/aboutaa/data_sales.html)> and Bartholomew's (巴爾多祿茂一世的) <<http://www.bartholomewmaps.com/>>。

購買這些資源也許不便宜，或是有版權限制，但是購買這些資料可以很快速的得到高品質的資料，而且不用自己去找這些資料。不論獲取數據資料的來源是什麼，有兩件事要記住，一是要注意原始資料的限制和是否適合研究的目的；二是任何資料在使用時都有著作權或其他限制，若是誤用，可能導致民事或刑事責任，並有可能導致其他人和機構不太願意繼續分享數據。

### 3.10 結論

取得空間資料通常需要花費時間，昂貴或兩者兼而有之，要想輕鬆取得並不容易。但是無論如何一旦空間資料變成可用的數位化型式，就變成與當初擷取它的原因更有力的使用資源。重要的是要記住全部數字化空間的原始數據仍有其局限性，無論是數據採集過程所提出的地圖或其他來源，還有許多額外的局限和錯誤。更多有關描述會在第 9 章提到，在資料獲取過程，要讓使用者能選擇適合的資料，良好的文件儲存和基本數據資料是必不可少的組成部份，但也只有符合使用者用途、需求之數據，才能確保會被廣泛採用。