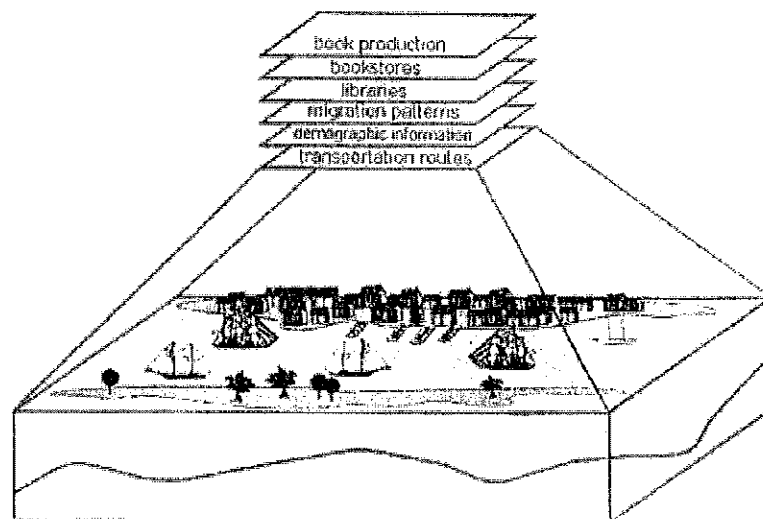


章節 2: 透過 GIS 看世界

2.1 介紹

這個章節給 GIS 軟體包使用存放他們的資料資料模型的一個詳細的描述。少數說這麼說在這兒全然是歷史的; 但是, 重要的是它呈現出一個簡要概要方式的來用 GIS 塑造世界。這一章節也建立了使用在這本書的術語。這些術語在不同的作者和不同的套裝軟體下會有差異。GIS 有很多適合歷史學者使用的東西。在 1991 年, 有三位學者收集了 GIS 的大量文獻, 並把這些文獻編輯成兩冊, 這也成為 GIS 的標準參考資料。1999 年這本書第一次由 Longley、Goodchild、Maguire 和 Rhind (Longley 等 1999) 完全地被重寫的再版出版, 編輯。大多數「巨著的初版」, 當它出名時, 就會被可以利用網路。2001 年四位作者寫了一篇關於 GIS 的期刊論文, 裡面有更多更廣的介紹。其他不錯的文章包括 Chairman, Demers, Heywood et al. and Martin。

像是第一章所說的, GIS 使用的傳統資料, 稱作屬性資料, 有時也被稱為非空間資料。每一筆屬性資料都有一個同類別的空間參考資料, 這些稱為空間資料。GIS 資料模型有兩種: 一種是向量式, 向量式的空間資料是由點、線、多邊形所組成。另一種是 raster, 研究區域通常被切割成許多像素。向量式模型空間式分離的, 而 raster 模型試圖描述空間為一連續的表面。也有別的方法把世界劃分成平面。像是不規則三角形法, 通常用來代表等高線, 穿過數位地形模型(DTMs, 亦稱數字式海拔模型或 DEMs)。



[Abstracting the real world into layers for the study of book history]

提取現實世界入層數為書歷史的研究

圖二之一：最上層的圖：書本製造。第二層：書局。第三層：圖書查詢系統。第四層：移民。第五層：人口統計資訊。第六層：運輸路線。

G I S並不是把研究區域裡，所有的資訊都存在單一個結構裡面；而是把它分成好幾個不同的圖層。每一個圖層都代表著不同類型的資訊。大部分G I S的文字，都像是在一個圖層裡的地景、道路、水系等等諸如此類。有兩個學者，他們對19世紀的印刷文化發展歷程有興趣，他們覺得書本的普及，原因跟書本本身的因素有關，包括一些資源的出現，像是印刷機、圖書查詢系統、書店。另外，書本普及的原因，還有像是運輸、人口遷移等等的原因。圖二之一顯示他們怎麼把真實世界轉成G I S的圖層。

我們可以透過來自全世界各地的空間資料和屬性數據去瞭解地理資訊系統模式的運作，空間資料可以彌補一小部分未經校正的原始圖像數據，並且屬性資料在不同的主題圖層下能夠聚合在一起使用是瞭解地理資訊系統全球化運作的重要關鍵。基本模式從根本上界定了地理資訊系統有什麼可以和什麼不可以做的，以及所有運作處理的長處和短處。另外還有一種採用以物件為導向去建構全球化地理資訊系統運作模式的方法。該方法的特點是不再細分資料層級，而是改以大目標分類成階級或等級作為替代方法。儘管上述做法有理論上的優勢，但仍很少有商用地理資訊系統封包軟體全面地推行這類做法。因此基於這個理由，以物件為導向的建構模式方法將不會在本指南手冊中討論。

2.2 屬性資料

屬性資料亦即每筆數據資料在形式上均可容易被大多數人清楚了解。許多地理資訊系統封包軟體雖囊括有自己的屬性資料庫，但也允許用戶可以建立參考連結到外部資料庫管理系統（DBMS）或電子分頁表格中。例如GIS的MapInfo，可以連結到Microsoft Excel的資料頁，ArcView連結到Dbase，或者ArcInfo連結到Oracle。屬性資料頻繁地使用統計或原文格式，然而事實上它們可以被使用中的資料庫管理系統存放成任何所支持的資料形式。越來越多地上述資料庫屬性資料包含了圖像格式、動畫、超連結格式、多媒體數化等等。

更多的資料庫管理系統（DBMS）被連結使用並成為GIS的關聯資料庫管理系統。這意謂著，有兩個或更多的資料庫平台可以基於一個共同知識我們通稱為關鍵詞的思考模式進行併合。例如歷史性資料數據往往不是一個地方的名字或身份證號碼（註記：地名並不被視作為有效空間資料，要得到一筆有效的空間資料，必須要有一個定位座標），且提供歷史性的資料數據是從各種多面向來源取得而

彙整成更完善，不是單靠空間資料就能界定的。再舉例來說明：有一個缺乏法效性資料組織的濟貧工會平台使用者，(那是一種早在 19 和 20 世紀初，在英格蘭和威爾斯普遍的窮人救濟相關行政管理單位)，一個議會選區投票統計組織，一些城鎮基礎的就業統計資料，我們將上述三個平台全部併合成一個具有相關性的連結，並連同所有關於"布里斯托(Bristol)"的數據資料，以此例來說，執行資料分析後將會出現一個單列。

這時有三個主要的問題須思考：第一，加入"無通則性"的連結，如實體查詢")，在不同平台上可能會得到不同結果。再者，重複的名稱也會發生問題，如"維特徹奇"(Whitchurch)這個地名在罕布夏郡(Hampshire)和什羅普郡

(Shropshire)都有出現。不同的軟體在處理這個問題上有不同的方式，最普遍的(理論上可行的)方法是複製資料的排列方式。使用一個以上的關鍵字欄位，如地名和國家，是其中一個方式。第三個問題是關於地名的使用，他們的拼法應該完全相同以產生吻合。在使用連字符號及撇號上即使是很小的差別，都會造成無法吻合。但如果使用標準化各種可能拼法的地名辭典並從權威的清單中建立單一拼法，或者使用身分證號，這樣的話資料分析則是能辦到的。不過建立這些資料庫是很耗時的。

許多屬性資料庫使用構造查詢語言(SQL) 允許靈活詢問和加入。這經常被實施雖然一個圖形用戶界面但跟隨基本的結構：

```
select <column names>   選擇專欄名字
from <table names>      目錄名
where <condition>       情況
```

如此，例如，我們有二張目錄：包含關於失業率的資料的`unemp'，並且包含關於嬰兒死亡率的資料的`inf_mort'。這些有以下領域：

字段名

區域名	區域名
區域的失業率	區域的嬰兒死亡率
區域的總人口	誕生的數量在區域

Unemp	
Field Name	Contents
Name	The area's place name
U_rate	The area's unemployment rate
Tot_pop	The area's total population

Inf_mort	
Field Name	Contents
Name	The area's place name
Im_rate	The area's infant mortality rate
Births	The number of births in the area

Table 2.1: Sample tables of attribute data

表 2.1: 屬性資料樣品目錄

The SQL query:

```
select name, u_rate, tot_pop
from unemp
where u_rate>10.0
```

像被顯示在表 2.2 將選擇名字、失業率, 和總人口從目錄 unemp 為地方以失業率的 10% 。

Name	U_rate	Tot_pop
Bolton	13.2	10,000
Oldham	12.1	7,500
Rochdale	10.9	8,000

Table 2.2: Sample data returned by the query above.

表 2.2: 樣品資料上面返回了由詢問。

關係加入並且被實施這樣。例如詢問:

```
select unemp.name, unemp.u_rate, inf_mort.im_rate
from unemp, inf_mort
where unemp.name=inf_mort.name
```

將選擇失業率從 unemp 和嬰兒死亡率從價值在名字字段在兩張目錄裡是相同的 inf_mort 。

Unemp.name	Unemp. U rate	Inf_mort. im rate
Bolton	13.2	120
Oldham	12.1	115
Rochdale	10.9	106
Burnley	9.4	98
Colne	9.1	89

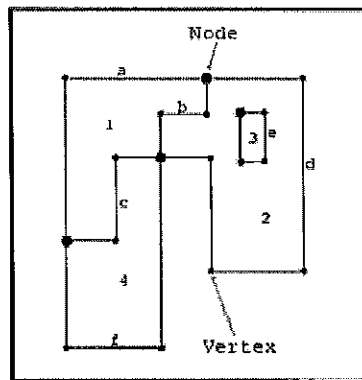
Table 2.3: Sample data returned by the query above

表 2.2: 樣品資料上面返回了由詢問(以關鍵字或簡單語句,向資料庫索閱資料)。

當關係資料庫和 SQL 對 GIS 的理解不是根本的, 知識他們可能是有用爲了提高對 GIS 資料和 GIS 軟體的理解。許多指南對於對 SQL 和關係資料庫的用途是可利用的: 看參考書目爲詳細資訊。

2.3 導向系統

在導向系統空間資料由點, 線代表或者, 或多角形點代表使用同等的對。(或弧或段) 的線由同等的對串代表給線和所有點座標的開始和終點線形變化方向。線的起動和末端點經常指結。多角形由完全地附寄創造區域使用一個或更多線。一個基本的多角形資料模型被顯示在表 2.2。連接的線方法創造這個模型爲人所知作爲拓撲結構。多角形和線需要拓撲學資訊。多角形需要知道身份證數字(或信件在圖 2.2 情況下組成它的界限的) 所有線段, 當線段通常知道哪個多角形是在它的左邊和它的右邊。例如在表 2.2, 多角形 1 由線段 a, b 一定並且 c 當多角形 3 由線段 e. 線段 c 只一定有多角形 1 到它的左邊和多角形 4 在它的右邊。



Polygons - line segments

1: a,b,c

2: b,d,e

3: e

4: c,f

Lines - co-ordinates

a: (1.36,4.38) (1.36,8.62) (5.12,8.6

b: (5.12,8.62) (5.12,7.65) (4.03,7.6

...

f: (1.36,4.38) (1.36,1.55) (4.03,1.5

[Simplified structure of polygons in a vector GIS]

在導向系統被多角形簡化的結構 GIS

如果多邊形能被使用，那麼是需要拓撲學的，就如同沒有其邊界必須被儲存兩次一般，舉例來說，線段 b 能被儲存為多邊形 1 與多邊形 2 的邊界部分，去記錄多邊形們不能套疊，以及研究區中每個位置只能被包含於一個多邊形內等等的外部資料模式是有重要性的，拓撲學也常被使用於將一堆蒐集的線段轉化為網路，經常基於每個節點認知到與其連接的線段們對其進行連接，所以換句話說，節點就是交叉點，當其作為運輸或被使用於其他網路時是非常實用的，比如計算兩節點間的最短距離。

Spatial data:

Attributed data:

Spatial data:		Attributed data:																					
		<table border="1"> <thead> <tr> <th>Polygon id</th><th>Parcel no.</th><th>Owner</th><th>Value</th></tr> </thead> <tbody> <tr> <td>1</td><td>437</td><td>A. Elliot</td><td>2201</td></tr> <tr> <td>2</td><td>617</td><td>J. Sutcliffe</td><td>1432</td></tr> <tr> <td>3</td><td>133</td><td>A. Elliot</td><td>5025</td></tr> <tr> <td>4</td><td>991</td><td>A. Smith</td><td>1692</td></tr> </tbody> </table>		Polygon id	Parcel no.	Owner	Value	1	437	A. Elliot	2201	2	617	J. Sutcliffe	1432	3	133	A. Elliot	5025	4	991	A. Smith	1692
Polygon id	Parcel no.	Owner	Value																				
1	437	A. Elliot	2201																				
2	617	J. Sutcliffe	1432																				
3	133	A. Elliot	5025																				
4	991	A. Smith	1692																				

[Linking spatial and attribute data]

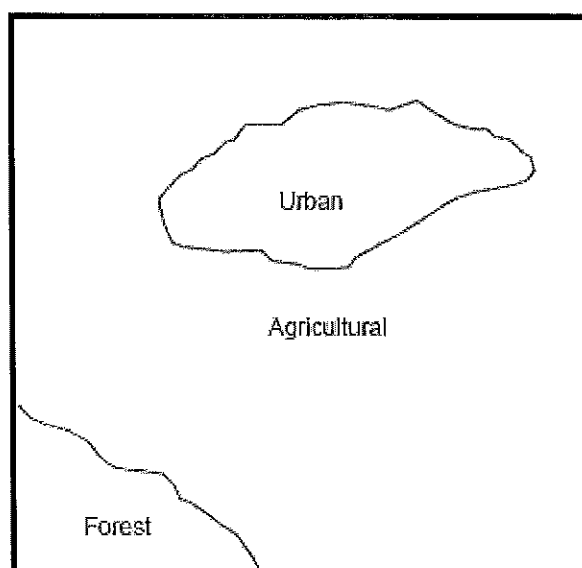
屬性資料以使用具關聯性的識別數值，與點、線與多邊形作連結，如圖 2.3 所示，在現代化的軟體中，拓樸學所創造與維持的那關聯的識別碼數值與介於空間資料與屬性資料的連結，皆對使用者隱藏起來，然而，對這些資料是如何集結而成的過程，做出初步理解是很重要的。

一些如城鎮或建築物的地貌，能被以多邊形或節點的方式呈現，選擇何種呈現方式取決於使用者的目地，由其是使用的比例尺，如果必要，多邊形能以節點來替代，此情形通常是呈現多邊形的中心點，被稱為距心的地方來呈現。

英國歷史 GIS (Gregory et al. 2002; Gregory and Southall 2000), SECOS (Gatley and Ell 2000), 比利時國土組織歷史 GIS (De Moor and Wiedemann 2001) 與美國國家歷史 GIS 均擁有包含豐富人口調查及其他資料的屬性資料庫，這些資料用以連結多邊形來代表行政的單位，利用於公布數據，如英國與威爾斯的註冊區與美國的州郡數，此時資料與地圖是合一的。

2.4 Raster systems 網格系統

網格資料模式把空間再細分成正方形的畫素或是其他固定的棋盤形，是爲了要提供研究區域一個連續性的圖像，而不是在細分成分離的點、線，和面。舉個例子來說，爲了要呈現地形，每個畫素都必須要有它的海拔高度作爲他的屬性；要呈現土地利用，每個畫素要有土地利用的分類依附它，等等。即使這對歷史學家不是那麼有用，衛星影像又是更複雜的網格資料。在影像上面，地球表面被分爲許多畫素，每個畫素都儲存相當的資訊與各種地表反射的光線，舉個例子來說，有多少綠色，多少藍色，多少紅色和多少紅外線。



1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	2	2	2	2	1	1	1	1
1	1	1	1	1	2	2	2	2	2	2	2	2	1	1
1	1	1	1	2	2	2	2	2	2	2	2	2	1	1
1	1	1	1	2	2	2	2	2	2	1	1	1	1	1
1	1	1	1	1	1	1	2	2	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	3	1	1	1	1	1	1	1	1	1	1	1	1	1
3	3	3	3	1	1	1	1	1	1	1	1	1	1	1
3	3	3	3	3	1	1	1	1	1	1	1	1	1	1
3	3	3	3	3	3	1	1	1	1	1	1	1	1	1

[The raster data model]

最簡單的網格資料格式為一個二度空間的陣列，每個畫素都有相符的值在格子裡，使這個陣列是立體的。首部資訊提供數據，舉個例子來說，畫素大小與左邊下半部角落在格子上的地點，當檔案的餘數簡單的構成每個細胞的值。例：Figure 2.4。原本的地圖呈現出三種土地利用，農地，都市和森林。網格將研究區域分成許多像素，每個像素分別配置數值以表現他的土地利用分類。像素配置的數值為覆蓋大部分的區域的類型。很明顯的，畫素的大小對模型來說是很重要的。太大會讓特徵圖像貧乏，太小會使得檔案使用不便。有許多高度發展的方法，

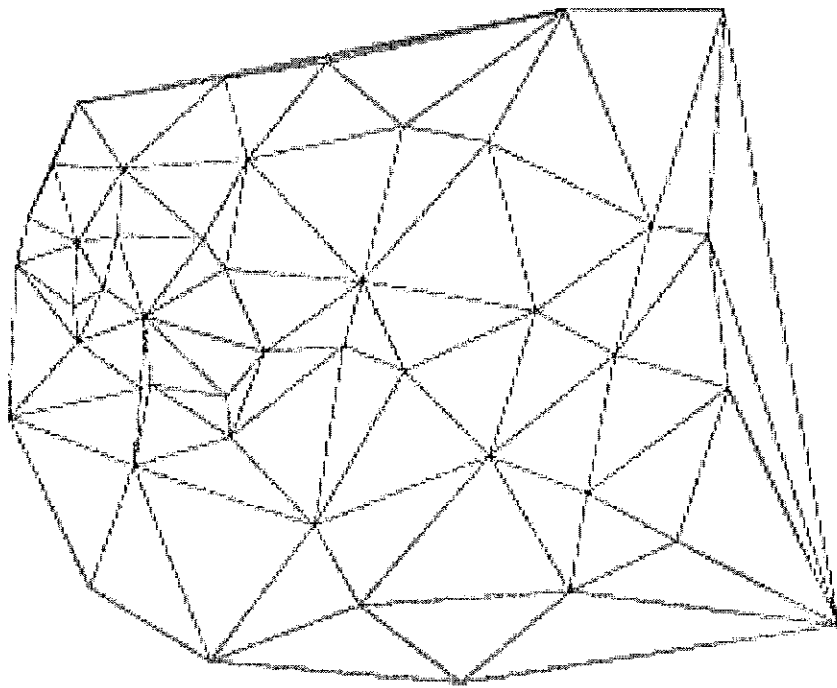
像紀錄重複次數和四元樹，可以用來壓縮檔案大小，但基本的模型還是一樣的。如果需要更詳盡關於網格架構的資訊，可以看第三章。〔不規則三角測量網絡 (TIN)〕

其中使用 DTMs 在歷史研究上是為了再現歷史景觀。這麼做是基於多個原因，包括探討過去人們如何出現(Harris 2002)，以調查研究戰役中的可見性和決策過程重塑歷史戰場 (Knowles,2004)，或是帶來生活舊地圖 (Rumsey and Williams 2002)。通常，當向量數據更適合於人類活動時，網格式資料更適合環境的應用。網格式系統模式複雜的空間模式帶著有限的屬性，例如土地使用模式，這是為人所熟知的，然而，儘管向量資料是更好的由於更明確清晰帶著複雜的屬性例如人口普查資料。這是個例外：馬丁 (1996b) 使用得到的網格式資料對 1981 和 1991 的人口普查資料做模型而且比較這兩者間的改變，聲稱網格式模式提供一個人口分佈區域為基礎更逼真的模式。在歷史背景裡一個好的網格式系統例子是由巴特萊和坎貝拉(1997)所提供。他們檢查十四世紀的驗屍調查而且使用這些來創造中世紀土地利用網格式 GIS，他們宣稱，這是最有潛力成為最複雜的調查直到十九世紀小部分的調查。一個網格式系統是可以被使用的，因為它提供一個土地面積完整的報導，它比多邊形系統提供更逼真的土地利用表述，因為它比多邊形系統更好處理不正確的原始資料。

2.5 Other systems: terrain modelling with TINs

其他系統：用 TINs 來塑造地形

一個不規則的三角形網路 (TIN) 可以從從點的資料中被創造出來，此外有 X 和 Y 的座標數值，也有通常代表高度的 Z 數值。地點的高度是以這種形式最普通的資料來源。產生 TIN 的第一部階段是繪製所有點的 X，Y 座標。下一步一條直線在每一點之間被畫出和它最近的鄰點。這就是圖 2.5 所顯示的。當所有點的高度是被知道的，這些線的坡度也可被知道。



[A Triangulated Irregular Network (TIN)]

如果有足夠的點這樣資料是有效的，這些可用來產生令人驚訝實際的、表面上的立體景色形狀表現，這就是大家所知道的數字化地形模型（DTM）。網格式和向量式資料裝飾數字化地形模型來提供地景上的面貌。這種模式經常出現在國際衛星傳送的國際電視節目消息上，例如，塞拉耶佛的圍攻定期以這種方式被描繪出來。

2.6 以「層」的方式帶來一切

在 GIS 地理系統中，信息中每個單獨的主題代表一層。普遍使用在建立一個研究領域有實際代表性的各種層次中，例如，一個數字地面模型（ α DTM）可以被用來塑造一個地區輪廓，一個柵格表面模型它的其土地使用，而一個點層用來代表建築物的重要性，一條線層用以代表河流，網絡層代表交通運輸系統，並且多角形分層堆積，代表實地格局和行政界限。這些都包含了相關的屬性。究竟何時使用一個單獨的層往往成為選擇性的問題。舉例來說，如果我們有三種類型的建築—教堂、醫院、博物館，這些可以儲存成 3 個獨立的層，又或者要說每個點代表什麼類型的建築，則可以儲存在一個單一屬性的層。最常見的 GIS 地理資訊系統軟件包，可以處理這些大部分類型的數據，雖然大多數將集中於一個。MapInfo 和 ArcGIS 都是主要載體系統，擁有有限柵格和地形建模功能。反之，跨度和 idrisi 方面主要是建立在柵格基礎上。

在學術領育域上，透過 GIS 研究某一區域的一個典範，是由 Pearson 和

Collier 提供在紐波特的一個在 Pembrokeshire 的地方行政區域所進行的一項任務(Pearson and Collier 1998)。他們對第 19 世紀中葉的土地的所有制和農業生產力感興趣。爲了調查這個主題，他們需要結合環境資料譬如土壤，斜率、面向和高度，此外還需以歷史統計方式，包括在早期的 19 世紀期間，在英國和威爾士執行的所有土地所有制和土地利用的詳細和全面的勘測，譬如人口調查資料，以及租稅的調查資料。租稅調查資訊被逐一記錄成多角形。各個領域所顯現出的多角形其中則包含各種屬性資料：所有者，居住者，字段名，耕種狀態，土地面積以及租稅金。首先，環境的數據，是依照每十公里的區域當中，以高度爲 50m 點，單獨的覆蓋物一個規則柵格。這也就是所創造出有高度的網格資料，並且需計算出斜率和面向的地形模型。像這樣集結所有的資料的方式，使得在 19 個世紀中葉影響農業生產力各種各樣的因素詳細的重建，並且開創一個更加精密的分析。

第二個例子由 Siebert (2000) 提供。他致力於重建東京的發展的歷史，並特別著重海岸線和流域、行政區域、人口分佈以及路軌網路等物理特點上。爲了這項研究，他爲各類型的結構要素創造各種各樣的圖層：線圖層代表了鐵路網路，或使用線(小河)或多角形(爲更大的那些流域)代表了海岸線和河川流域；同時，以多角形清楚地代表了行政區域，也就是人口調查資料。以 Pearson 和 Collier 的任務來看，用這樣的方法進行研究，建立了嶄新又全面的研究領域。

2.7 Conclusions 結論

在 GIS 後面的那些核心想法那個，試圖描述地球表面的特徵。在做過程中的第一個階段就是帶作爲層對研究和抽象他們相關的現實世界主題，帶有每個主題被描述透過單一的層次。有特色在每層上描述使用一空間的結合並且把歸於與每個層次的空間數據由 4 類圖表原語之一組成一起的數據：點，線，多角形或者像素。導因的數據被因此非常從現實世界提取。數據可能被有效地描述爲使用這個模型，因爲處理的模型的獨特能力，GIS 提供一件強有力的研究工具但是，這個架構基本上限制路，GIS 可以被使用和特徵不能被以這種方法有效地描述的地方，使用 GIS 可能不是合適的。

大多數已經被在這章裡描述的作法被現代 GIS 軟體自動執行並且被很好地躲過用戶的視線，但是理解 GIS 代表世界的模式是重要的。很多數據集完全不適合這模式。這不可能值得把很多努力投入 GIS。這裡他們做，這個模型代表一種把空間併入一個分析或者一個特別的位置的探索的強有力的模式，它是一塊的只幾平方米，一整個國家或者甚至全世界。除描述空間，很多其他問題的原始方法之外也需要被記住。這些包括規模，準確和不確定性。GIS 研究團契用掉擔心的大數量的精力，GIS 用戶將不理解這些限制。不過，爲史學家所熟知，他們是正對原先的來源的限制透過把來源轉變成數字化的形式建立的其他限

制。問題適合空間引用這些更錯綜複雜以致於他們可能乍看起來出現的數據。有關問題將在隨後的章裡被更詳細討論。