

ĐẠI HỌC HUẾ  
TRƯỜNG ĐẠI HỌC KHOA HỌC  
KHOA MÔI TRƯỜNG



**BÀI GIẢNG**

**HỆ THỐNG THÔNG  
TIN ĐỊA LÝ**

*(Sử dụng cho sinh viên Khoa Môi trường)*

**GEOGRAPHIC  
INFORMATION SYSTEMS  
(GIS)**

**LÊ BẢO TUẤN**

lebaotuan@gmail.com  
Khoa Môi trường  
Trường Đại học Khoa học Huế

# MỤC LỤC

## CHƯƠNG 1: Giới thiệu về GIS

- 1.1. Giới thiệu
  - 1.1.1. GIS là gì?
  - 1.1.2. Nguồn gốc và sự phát triển của GIS
  - 1.1.3. GIS: Một công cụ phổ biến
  - 1.1.4. Tại sao chúng ta cần GIS?
  - 1.1.5. GIS trong hoạt động thực tế
  - 1.1.6. Khoa học thông tin địa lý
- 1.2. Thành phần và chức năng của GIS
  - 1.4.1. Thành phần của GIS
  - 1.4.2. Chức năng của GIS
- 1.3. Một số ứng dụng GIS
- 1.4. Chi phí cho GIS

## CHƯƠNG 2: Các mô hình dữ liệu

- 2.1. Giới thiệu chung về dữ liệu
- 2.2. Giới thiệu chung về bản đồ
  - 2.2.1. Khái niệm bản đồ
  - 2.2.2. Các đối tượng bản đồ
  - 2.2.3. Các hệ quy chiếu và hệ tọa độ sử dụng ở Việt Nam
- 2.3. Cấu trúc dữ liệu
  - 2.3.1. Cấu trúc dữ liệu vector
  - 2.3.2. Cấu trúc dữ liệu raster
  - 2.3.3. Cấu trúc dữ liệu TIN
- 2.4. Cấu trúc cơ sở dữ liệu
  - 2.4.1. Cơ sở dữ liệu không gian
  - 2.4.2. Cơ sở dữ liệu phi không gian
  - 2.4.3. Kết nối các đối tượng và thuộc tính

## CHƯƠNG 3: Nhập dữ liệu

- 3.1. Khái quát
- 3.2. Các công nghệ thu thập dữ liệu
  - 3.2.1. Số hóa
  - 3.2.2. Quét bản đồ
  - 3.2.3. Viễn thám
  - 3.2.4. Đo đạc
  - 3.2.5. Hệ định vị toàn cầu (GPS)
  - 3.2.6. Chuyển đổi dữ liệu
- 3.3. So sánh và lựa chọn phương pháp nhập dữ liệu

## **CHƯƠNG 4: Quản lý dữ liệu**

- 4.1. Nguyên tắc quản lý dữ liệu
- 4.2. Quản lý dữ liệu không gian
- 4.3. Quản lý dữ liệu thuộc tính

## **CHƯƠNG 5: Phân tích dữ liệu**

- 5.1. Giới thiệu chung
- 5.2. Các phép phân tích dữ liệu cơ bản
  - 5.2.1. Lựa chọn và phân loại
  - 5.2.2. Phân lập
  - 5.2.3. Tạo vùng đệm (buffer)
  - 5.2.4. Chồng ghép
  - 5.2.5. Phân tích mạng
- 5.3. Quy trình phân tích địa lý

## **CHƯƠNG 6: Hiển thị và xuất dữ liệu**

- 6.1. Mở đầu
  - 6.1.1. Màn hình
  - 6.1.2. Máy in
  - 6.1.3. Máy vẽ
- 6.2. Hiển thị dữ liệu
  - 6.2.1. Hiển thị bản đồ và bảng
  - 6.2.2. Hiển thị bản đồ vector và raster
- 6.3. Xuất dữ liệu dưới dạng bản đồ
- 6.4. Chuẩn dữ liệu và chất lượng dữ liệu
  - 6.4.1. Vấn đề chuẩn dữ liệu
  - 6.4.2. Các yếu tố chất lượng dữ liệu
  - 6.2.3. Các nguồn sai sót
  - 6.2.4. Kiểm tra và sử dụng dữ liệu

## **CHƯƠNG 7: Các phát triển mới trong GIS**

- 7.1. GPS - Bản đồ di động
- 7.2. Cải thiện công nghệ viễn thám
- 7.3. Bản đồ Internet
- 7.4. Thế giới ảo
- 7.5. GIS mở

*Tài liệu tham khảo*

# CHƯƠNG 1

## Giới thiệu về GIS

### 1.1. Giới thiệu

Địa lý luôn có vai trò quan trọng đối với con người. Trong thời kỳ đồ đá, những người đi săn đã đoán trước vị trí của những con mồi, những người thám hiểm đầu tiên trên thế giới cũng đã sống và chết bằng chính kiến thức về địa lý của họ và xã hội ngày nay đang sống và làm việc dựa vào sự hiểu biết về vấn đề ai thuộc vị trí nào. Trong các mẫu bản đồ và thông tin không gian, địa lý ứng dụng đã và đang phục vụ cho sự khám phá, quy hoạch, hợp tác cũng như xung đột trong suốt 3000 năm qua và bản đồ là một trong số các tài liệu đẹp nhất trong nền văn minh của chúng ta.



*Hình 1.1: Một bản đồ Châu Mỹ năm 1640 của tác giả Hondius*

Đa số các kiến thức về địa lý của chúng ta được áp dụng và các công việc hàng ngày; ví dụ như khi chúng ta đổ nhau về các tuyến đường của một khu vực được coi là mê cung đường phố ở một nơi nào đó hay khi chúng ta tìm kiếm một trạm xăng gần nhất nào đó. Thông tin không gian có một sự tác động mạnh mẽ hơn đối với cuộc sống của chúng ta, thường trong một phạm vi nào đó mà ta không nhận thấy để giúp chúng ta sản xuất ra lương thực, năng lượng, trang phục mà chúng ta sử dụng hàng ngày và kể cả những trò giải trí mà chúng ta tận hưởng.

Bởi lẽ thông tin không gian là rất quan trọng nên chúng ta đã và đang phát triển các công cụ được gọi là hệ thống thông tin địa lý (GIS - Geographic Information Systems) để trợ giúp ta cùng với các kiến thức địa lý mà ta có được. Một hệ thống thông tin địa lý (GIS) giúp chúng ta tập hợp và sử dụng các dữ liệu không gian (ở đây, cụm từ viết tắt GIS được dùng để chỉ một hay nhiều hệ thống thông tin địa lý). Một số thành phần GIS hoàn toàn thuộc về mặt kỹ thuật; chúng bao gồm các thiết bị thu thập dữ liệu rất hiện đại, các mạng lưới giao tiếp tiên bộ và tin học phức tạp. Các phương pháp của GIS khác thì lại rất đơn giản, chẳng hạn như khi một cây bút chì và một mẫu giấy được sử dụng để xác nhận ngoài thực địa một bản đồ.

Như nhiều khía cạnh của cuộc sống trong 5 thập kỷ trở lại đây, cách thức chúng ta tập hợp và sử dụng các dữ liệu không gian đã và đang được thay đổi một cách sâu sắc bởi các công nghệ điện tử hiện đại và các phần cứng, phần mềm GIS là kết quả tất yếu quan trọng của sự phát triển công nghệ đó. Việc thu và xử lý các dữ liệu không gian được đẩy mạnh trong vòng 3 thập kỷ gần đây và vẫn đang tiếp tục tiến triển.

Chìa khóa của tất cả các định nghĩa của GIS là “cái gì” và “ở đâu”. GIS và phân tích không gian có liên quan đến vị trí chính xác và tương đối của các đặc trưng cũng như các đặc điểm, thuộc tính của các đặc trưng đó. Vị trí của các đối tượng không gian quan trọng như các dòng sông, suối có thể được ghi lại, kể cả các thông tin có liên quan như kích thước, tốc độ dòng chảy, chất lượng nước hay các loài các được tìm thấy trong sông, suối đó. Quả thật, các thông tin thuộc tính đó thường phụ thuộc vào sự sắp xếp mức độ quan trọng của các đối tượng khác nhau. GIS có khả năng hỗ trợ cho việc phân tích và hiển thị các mối quan hệ không gian đó.

### **1.1.1. GIS là gì?**

Một hệ thống thông tin địa lý là một công cụ cho việc tạo ra và sử dụng thông tin không gian. Tuy nhiên, hiện nay có nhiều định nghĩa, quan niệm hay cách nhìn nhận và cách hiểu khác nhau về GIS do GIS là một công nghệ mới phát triển nhanh, có ứng dụng trong nhiều lĩnh vực hoạt động của con người; với mục đích thực hành, GIS có thể được định nghĩa như sau:

*một hệ thống dựa vào máy tính để trợ giúp cho công tác thu thập, bảo vệ, lưu trữ, phân tích, xuất và phân phối dữ liệu và thông tin không gian.*

Khi được sử dụng một cách sáng suốt thì GIS có thể giúp cho ta sống khỏe mạnh hơn, giàu có hơn và an toàn hơn.

Mỗi người sử dụng GIS có thể quyết định đặc tính nào là quan trọng và cái gì là quan trọng đối với chúng. Ví dụ như, rừng là quan trọng đối với nhiều người. Chúng bảo vệ nguồn nước của chúng ta, sản lượng gỗ, là nơi ẩn náu của thế giới hoang dã và cung cấp không gian để tái tạo. Tất cả chúng ta đều có liên quan đến mức độ của việc thu hoạch vụ mùa, việc sử dụng đất xung quanh ta, việc xả thải gây ô nhiễm của các khu công nghiệp lân cận hay việc cháy rừng xảy ra ở đâu và khi nào. Sự quản lý rừng có hiểu biết đòi hỏi kiến thức tối thiểu về tất cả các nhân tố có liên quan và có thể là cả sự phân bố về mặt không gian của các nhân tố đó. Các vùng đệm gần các con sông có thể bảo vệ nguồn nước, các khoảng rừng trống có thể ngăn ngừa sự lan rộng của lửa khi có cháy rừng và các tác nhân ô nhiễm phát tán ngược chiều gió có thể không gây thiệt hại cho các khu rừng như khi chúng phát tán theo hướng gió. Một hệ thống thông tin địa lý hỗ trợ tối đa trong việc phân tích các mối quan hệ không gian này và sự tác động qua lại giữa chúng. Một hệ thống thông tin địa lý cũng đặc biệt có ích cho việc hiển thị dữ liệu không gian và báo cáo kết quả của việc phân tích không gian. Trong nhiều trường hợp, GIS là cách duy nhất để giải quyết các vấn đề liên quan đến không gian.

### ***1.1.2. Nguồn gốc và sự phát triển của GIS***

GIS được khai phá vào những năm 1960 từ một sáng kiến bản đồ hóa công tác quản lý rừng của người Canada. GIS tiếp tục được phát triển thông qua việc tìm kiếm của các nhà nghiên cứu ở các trường đại học và chính phủ Canada, Mỹ và các quốc gia khác nhằm mục đích giới thiệu các yếu tố địa lý của Trái đất bằng cách sử dụng một hệ cơ sở dữ liệu máy tính, hiển thị nó trên thiết bị đầu cuối của máy tính và vẽ bản đồ ra giấy. Họ cũng đã phát triển các chương trình máy tính để tìm kiếm và phân tích các dữ liệu này một cách nhanh chóng. Đến những năm 1970, một số hợp tác đã được thiết lập để phát triển và bán các hệ thống máy tính sử dụng cho việc vẽ bản đồ và phân tích. Ngày nay, hai hãng phát triển phần mềm GIS hàng đầu đã tìm thấy nguồn gốc của họ trong những ngày đầu mới phát triển, tuy nhiên vào thời kỳ đó, họ đã chú trọng và đi sâu vào hai hướng công nghệ khác biệt nhau.

Tập đoàn Intergraph của Huntsville, Alabama đã tập trung vào dữ liệu đầu vào và khả năng lưu trữ có hiệu quả của dữ liệu GIS cũng như sự chuẩn bị các bản đồ được xuất ra từ máy tính để cạnh tranh với các bản đồ truyền thống về chất lượng vẽ bản đồ. Viện nghiên cứu hệ thống môi trường, ESRI (the Environmental Systems Research Institute) của Redlands, California tập trung vào việc cung cấp bộ công cụ lệnh máy tính để phân tích các dữ liệu GIS. Qua nhiều năm, cả hai công ty trên đã không ngừng phát triển khả năng riêng của các hệ thống của họ.

Ban đầu, chỉ các cơ quan chính phủ, các ngành phục vụ cộng đồng và các tập đoàn lớn mới có thể có khả năng sử dụng GIS bởi vì chi phí cao. Trên nền tảng các máy chủ và máy tính con, một trạm nghiên cứu GIS điển hình đòi hỏi một chi phí

hơn 100 ngàn đôla, bao gồm tất cả các phần cứng, phần mềm và công tác đào tạo. Tuy nhiên, thị trường GIS được mở rộng một cách mạnh mẽ vào đầu những năm 1980 nhờ vào các tạp chí thương mại, các hội nghị và các sự hợp tác chuyên nghiệp truyền bá cho toàn thế giới về các lợi ích của GIS. GIS phát triển nhanh như nắm cùng với sự xuất hiện của máy tính cá nhân và các phần mềm GIS đã nhanh chóng thích nghi với bước ngoặt mới và ít đắt đỏ hơn này. Và chi phí của các phần mềm đã giảm trong khi số lượng người sử dụng ngày càng gia tăng.

Mặc dù vậy, thời gian đầu, GIS được ít người sử dụng bởi nó đòi hỏi phải có phần mềm GIS được cài đặt trong máy tính và được đào tạo để sử dụng nó. Tuy nhiên, hàng chục triệu đôla đã được đầu tư vào xây dựng hệ cơ sở dữ liệu GIS từ các bản đồ giấy, ảnh không gian và ảnh vệ tinh. Các dữ liệu này vẫn chưa đạt được trên quy mô rộng lớn cho đến khi các chuyên gia GIS quyết định vẽ chúng ra trên giấy để tiến hành phân loại chúng. Đến những năm 1990, Internet đã mở cánh cửa cho việc đưa các dữ liệu GIS có giá trị này đến với người sử dụng trên toàn thế giới.

Ngày nay, có hàng trăm website đăng tải dữ liệu GIS trực tuyến trên mạng toàn cầu Internet. Bất kỳ một ai có thể sử dụng các trình duyệt web đều có thể truy cập và xem các dữ liệu GIS. Và như là một kết quả, thị trường các sản phẩm và dịch vụ GIS với lợi nhuận 7 tỉ đôla năm 1999 đang ngày càng gia tăng với tốc độ gần 13% mỗi năm.

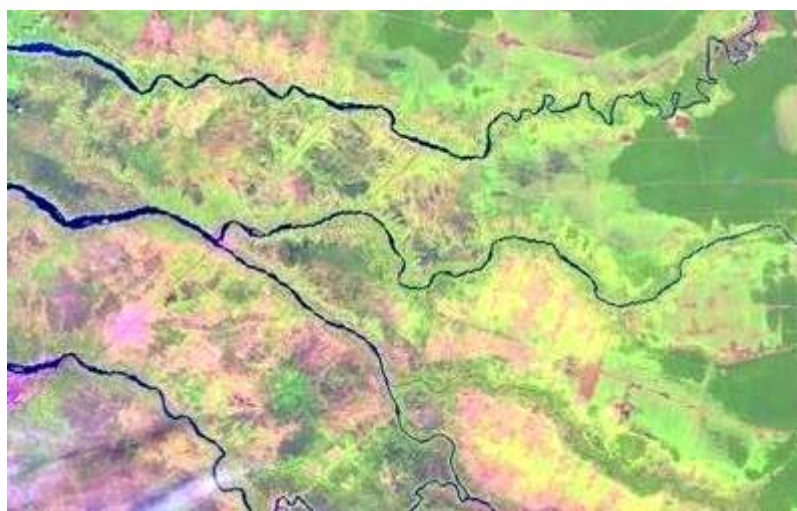
Hệ GIS điển hình được thiết lập trên một số khái niệm cơ bản sau:

- Các đặc điểm của thế giới thực trên bề mặt Trái đất được mô tả lại trên một hệ quy chiếu bản đồ và được lưu lại trong máy tính. Đồng thời, máy tính cũng lưu lại lưới chiếu và các thuộc tính của các đặc điểm bản đồ đó để có thể trả lời các câu hỏi như “chúng ở đâu?” và “chúng là cái gì?”.
- Các đặc điểm bản đồ có thể được hiển thị hoặc vẽ ra khi ta kết hợp bất kỳ hai hay nhiều đối tượng và hầu như trên bất kỳ một tỷ lệ bản đồ. Tin học hóa các dữ liệu bản đồ phải được sử dụng một cách linh hoạt hơn so với các bản đồ giấy truyền thống.
- GIS có khả năng phân tích các mối quan hệ trong không gian giữa các đặc điểm bản đồ.

### ***1.1.3. GIS: Một công cụ phổ biến***

Việc sử dụng GIS đã trở nên phổ biến và lan rộng trong vòng 2 thập kỷ qua. GIS được sử dụng trong nhiều lĩnh vực khác nhau, từ khảo cổ học cho đến động vật học và nhiều ứng dụng mới của GIS tiếp tục xuất hiện. GIS là những công cụ cần thiết trong kinh doanh, hoạt động của chính phủ, giáo dục và các tổ chức phi lợi nhuận; và việc sử dụng GIS đang trở nên có tính bắt buộc trong nhiều trường hợp. GIS được sử dụng chống lại tội ác, bảo vệ các loài bị nguy hiểm, giảm thiểu sự ô nhiễm, đối phó với các tai biến thiên nhiên, phân tích sự lan truyền bệnh dịch

AIDS và để cải thiện sức khỏe cộng đồng; tóm lại, GIS là phương tiện để giải quyết một số vấn đề xã hội cấp bách của chúng ta. Trong thực tế, các công cụ GIS giúp tiết kiệm được tổng số hàng tỉ đôla hàng năm trong việc phân phát các dịch vụ và hàng hóa thương mại và của chính phủ. Hiện nay, GIS giúp cho công tác quản lý hàng ngày các nguồn tài nguyên thiên nhiên và nhân tạo, bao gồm cả hệ thống công rãnh, nước, năng lượng và mạng lưới giao thông. GIS đang ở vị trí trái tim của một trong số các quy trình quan trọng nhất của nền dân chủ Mỹ, đó là quy trình xem xét lại định hướng phát triển được ủy nhiệm theo quy định của hiến pháp của Quốc hội Mỹ và từ đó liên quan đến việc phân phối những đồng đôla tiền thuế và các nguồn lực khác của chính phủ.



*Hình 1.2: Ảnh vệ tinh cung cấp thông tin về vùng phân bố rừng và nguồn nước*

#### **1.1.4. Tại sao chúng ta cần GIS?**

GIS là cần thiết phần nào đó bởi vì dân số trên thế giới đang tăng nhanh và công nghệ đang ở trình độ cao trong khi các tài nguyên, đặc biệt là không khí và đất đang ở trong tình trạng giới hạn do hoạt động của con người. Dân số thế giới đã tăng gấp đôi trong 50 năm qua và đạt mức 6 tỉ người và dường như chúng ta sẽ có thêm 5 tỉ người nữa trong vòng 50 năm tới. 100.000 năm đầu tiên của sự tồn tại của con người đã gây ra rất ít các tác động lên các tài nguyên của Thế giới trong khi chỉ 300 năm vừa qua, con người đã làm biến đổi thường xuyên đa phần bề mặt của Trái đất. Khí quyển và các đại dương đã cho thấy sự suy giảm khả năng hấp thu tốt các khí CO<sub>2</sub> và nitơ, hai sản phẩm xả thải chính của con người. Bùn đã “bóp nghẹt” nhiều con sông và có vô số các ví dụ khoanh vùng nơi mà ôzôn, hay các chất thải độc hại khác làm tổn hại sức khỏe cộng đồng. Đến cuối thế kỷ 20, đa số các vùng đất thích hợp đều có người ở và chỉ một vài phần trăm nhỏ diện tích đất bề mặt chưa bị trồng, chăn thả, cắt xén, xây dựng lên trên, hút nước, ngập lụt hay bị biến đổi theo cách khác bởi con người.

GIS giúp chúng ta xác định các vấn đề môi trường bằng cách cung cấp các thông tin chủ yếu về nơi mà các vấn đề đó xảy ra và ai bị tác động bởi chúng. GIS



giúp ta xác định nguồn, vị trí và quy mô của các tác động môi trường có hại, và có thể giúp ta đưa ra kế hoạch hành động để quan trắc, quản lý và giảm thiểu các thiệt hại môi trường.



**Hình 1.3:** Ảnh vệ tinh chụp cộng đồng ven biển vùng Banda Aceh, Indonesia ngày 23/06/2004 (trái) và ngày 28/12/2004 (phải) bị thiệt hại nặng nề sau khi xảy ra sóng thần

Các tác động của con người lên môi trường đã tạo nên áp lực về mặt xã hội và điều đó có vai trò thúc đẩy mạnh đối với việc thông qua GIS như là một giải pháp. Các xung đột trong vấn đề sử dụng tài nguyên, liên quan đến sự ô nhiễm và công tác đề phòng để bảo vệ sức khỏe cộng đồng đã dẫn đến yêu cầu giải quyết mang tính pháp lý và hoàn toàn phải có sự xem xét về mặt địa lý. Đạo luật các loài bị nguy hiểm của Mỹ (ESA - Endangered Species Act) năm 1973 là một ví dụ về tầm quan trọng của địa lý trong vấn đề quản lý tài nguyên. ESA yêu cầu sự bảo vệ thỏa đáng đối với các sinh vật quý hiếm và đang bị đe dọa. Sự bảo vệ hiệu quả đòi hỏi phải bản đồ hóa các nơi ở có thể có và phân tích các loài sinh vật và các kiểu di trú. Vị trí của các cá thể sống sót có liên quan đến vấn đề sử dụng đất hiện nay và trong tương lai của con người phải được phân tích và phải có các hoạt động được thực hiện để chắc chắn rằng các loài vẫn còn sống sót. GIS đã chứng tỏ là những công cụ rất hữu ích trong việc thực hiện tất cả các nhiệm vụ trên. Pháp luật cũng thúc đẩy việc sử dụng GIS trong các nỗ lực ở nhiều lĩnh vực khác như các dịch vụ khẩn cấp, phòng chống lũ lụt, theo dõi, quản lý thiên tai, quy hoạch và phát triển cơ sở hạ tầng.

Nhiều doanh nghiệp cần GIS vì họ cố gắng gia tăng hiệu quả trong việc phân phối sản phẩm và dịch vụ của mình. Các doanh nghiệp bán lẻ tiên hành đặt các cửa

hàng của mình dựa vào một số yếu tố không gian có liên quan như: các khách hàng tiềm năng phân bố ở đâu? Khu vực phân phối của các doanh nghiệp đối thủ là đâu? Đâu là các vị trí tiềm năng cho một cửa tiệm mới? Luồng giao thông nào gần với các cửa tiệm hiện nay và việc đỗ xe và mua sắm ở các cửa hàng đó có dễ dàng hay không? Việc phân tích không gian được sử dụng hàng ngày để trả lời các câu hỏi đó. GIS cũng còn được sử dụng trong hàng trăm ứng dụng khác trong kinh doanh như để vạch các tuyến đi của phương tiện phân phối sản phẩm, hướng dẫn cho việc quảng cáo, thiết kế các tòa nhà, lập kế hoạch xây dựng hay mua bán đất.

Các tổ chức cộng đồng cũng được sự hỗ trợ của GIS bởi lẽ GIS hỗ trợ các chức năng của chính phủ. Sự phát triển đô thị gây ra sự thay đổi về cảnh quan và GIS là một công cụ quan trọng cho việc quy hoạch hợp lý. Các phương tiện của dịch vụ khẩn cấp được thường xuyên điều động và việc vạch tuyến đường đi có sự trợ giúp của GIS. GIS dùng cho việc đáp lại các tình huống khẩn cấp được phát triển và cài đặt rộng rãi để phản ứng nhanh các yêu cầu khẩn. Người gọi điện đến số máy khẩn cấp được xác nhận tự động thông qua số điện thoại gọi đến. Và số điện thoại giúp cho việc xác nhận địa chỉ của tòa nhà đang có sự cố cũng như xác định cá trạm cứu hỏa, cảnh sát hay cấp cứu gần nhất. Một bản đồ đường đi ngay lập tức được thiết lập để cung cấp đường đi tối ưu đến nơi cần sự hỗ trợ và được gọi về các trạm ứng cứu phù hợp cùng với hệ thống báo động tự động.



**Hình 1.4:** Máy tính bỏ túi là một ví dụ của thiết bị ứng dụng công nghệ GIS

Trên đây là một số trường hợp mà sự hỗ trợ của GIS là rất cần thiết đối với xã hội loài người hiện nay. Tuy nhiên, sự phát triển của công nghệ lại tạo nên động lực thúc đẩy sự phát triển và ứng dụng của GIS. Việc phân tích không gian trở nên có ích hơn nhiều với máy tính tốc độ cao hơn và ổ đĩa cứng lưu trữ lớn hơn. Đồng thời việc giá cả thiết bị ngày càng giảm do cạnh tranh đã giúp cho GIS ngày càng phổ biến và phát huy tốt hơn sức mạnh của nó.

### **1.1.5. GIS trong hoạt động thực tế**

Việc tổ chức, phân tích và phân phối, chia sẻ dữ liệu không gian được ứng dụng rộng rãi nhằm mục đích cải thiện cuộc sống. Sau đây là một ví dụ minh họa cho việc GIS đã được sử dụng như thế nào.

Ông Marvin Matsumota sống sót nhờ vào GIS. Ông là một vận động viên đi bộ 60 tuổi, bị lạc trong vườn quốc gia Joshua, Mỹ rộng 300.000 ha sa mạc. Trước đó vài năm, khoảng 6-8 người đi bộ cũng đã lạc trong vườn và mất tích. Vì sự nguy hiểm, Cục vườn quốc gia Mỹ tổ chức tìm kiếm và giải cứu ông Marvin bằng

cách tuần tra bằng đường bộ, sử dụng ngựa và các phương tiện đi lại, và tìm kiếm bằng trực thăng. Cuộc tìm kiếm được tổ chức và hướng dẫn bằng việc sử dụng GIS. Trung tâm GIS đã tải dữ liệu bản đồ nền và các thuộc tính địa lý của khu vực vào các thiết bị. Đội tìm kiếm mang theo các thiết bị định vị khi đi sâu vào khu vực và ghi lại vị trí và tiến trình tìm kiếm. Dữ liệu được tải về trở lại trung tâm GIS để cập nhật liên tục dữ liệu. Người quản lý tiến hành đánh giá và phân vùng đã tìm kiếm và đưa ra các kế hoạch hành động phù hợp một cách kịp thời. Dữ liệu bản đồ chỉ rõ vùng đã được tìm kiếm và cách thức tìm kiếm được thực hiện... Marvin được tìm thấy vào ngày thứ 5, bị mất nước nhiều và chấn thương đầu. Đội tìm kiếm đã điện vị trí chính xác cho trực thăng cứu hộ để đưa Marvin đến bệnh viện. Các bác sĩ cho biết nếu chậm thêm 1 ngày nữa thì Marvin sẽ chết - 1 ngày được tiết kiệm nhờ sử dụng hiệu quả GIS. Sau một tuần ở bệnh viện và vài tháng nghỉ dưỡng tại nhà, ông Marvin đã hoàn toàn hồi phục trở lại.



### ***1.1.6. Khoa học thông tin địa lý***

Trong khi chúng ta định nghĩa GIS là hệ thống thông tin địa lý (Geographic Information Systems) thì còn có một GIS khác đó là khoa học thông tin địa lý (Geographic Information Science). Thông thường thì cụm từ GIS được sử dụng cho hệ thống thông tin địa lý, tuy nhiên, đôi lúc GIS cũng được dùng để chỉ khoa học thông tin địa lý (GIScience). Sự khác biệt là rất quan trọng vì tương lai phát triển của GIS phụ thuộc vào quá trình trong khoa học thông tin địa lý (GIScience). GIScience rộng lớn hơn nhiều so với GIS vì GIScience được hình thành trên cơ sở

lý thuyết mà GIS dựa vào. GIScience bao gồm các vấn đề liên quan của GIS nhưng đồng thời cũng cố định nghĩa lại các khái niệm trong địa lý và thông tin địa lý trong hoàn cảnh của thời đại số. GIScience có liên quan đến các vấn đề như chúng ta khái niệm hóa địa lý như thế nào và chúng ta thu thập, thể hiện lại, lưu trữ, hình dung, phân tích, sử dụng và giới thiệu các khái niệm địa lý như thế nào. GIScience nghiên cứu kỹ không chỉ các câu hỏi kỹ thuật cho các nhà địa lý ứng dụng, nhà kinh doanh, nhà quy hoạch, nhân viên an toàn cộng đồng và các đối tượng khác mà GIScience cũng có hướng đi vào các vấn đề cơ bản hơn như chúng ta hiểu về không gian như thế nào? Bằng cách nào chúng ta có thể giới thiệu về các khái niệm không gian một cách tốt nhất? Tâm lý con người giúp đỡ hay cản trở việc tranh luận về không gian có hiệu quả như thế nào?

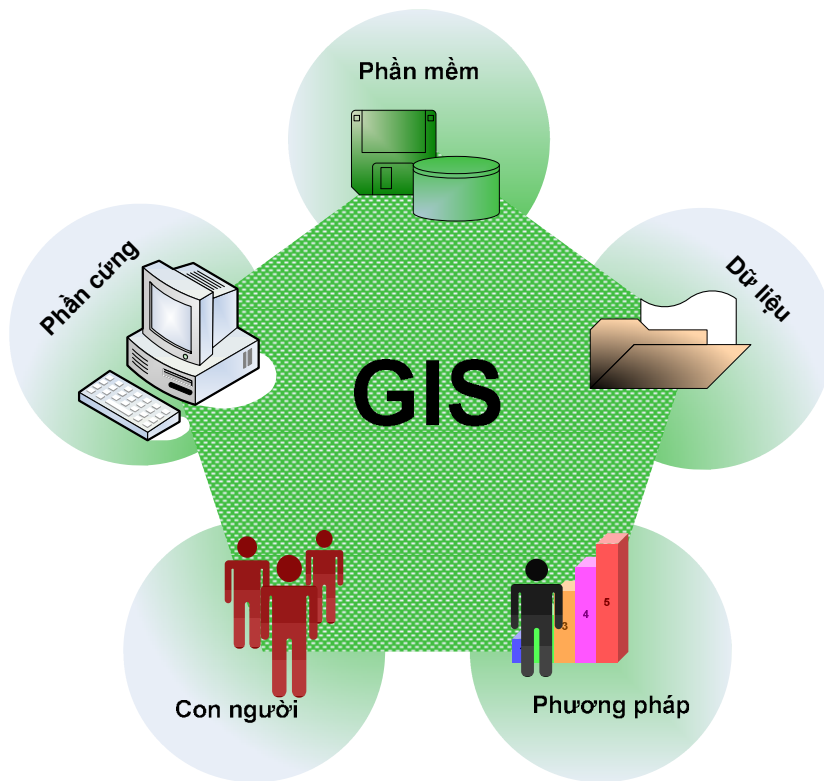
Khoa học được mô tả như một sự trợ giúp của kỹ thuật trong thế giới ứng dụng. Một sự suy luận hay hơn, tốt hơn, tối ưu hơn có thể là cha đẻ của một kỹ thuật. Định nghĩa một cách tỉ mỉ thì GIS có tính kỹ thuật cao hơn so với khoa học. Từ khi GIS trở thành công cụ chúng ta sử dụng để giải quyết vấn đề, chúng ta sẽ sai lầm nếu nghĩ rằng nó là điểm khởi đầu và điểm kết thúc trong lập luận về mặt địa lý. Và sự hiểu biết về GIScience mang tính quyết định đối với sự phát triển xa hơn của GIS và trong một số trường hợp sẽ mang tính quyết định đối với việc ứng dụng hiệu quả GIS.

## **1.2. Thành phần và chức năng của GIS**

### ***1.2.1. Thành phần của GIS***

Một hệ thống thông tin địa lý bao gồm phần cứng, phần mềm, dữ liệu, con người và bộ quy định ở cấp độ tổ chức. Các hợp phần này phải được hợp nhất tốt để phục vụ cho việc sử dụng GIS hiệu quả; và sự phát triển và tương thích của các hợp phần là một quá trình lặp đi lặp lại theo chiều hướng phát triển liên tục. Việc lựa chọn và trang bị phần cứng và phần mềm thường là những bước dễ dàng nhất và nhanh nhất trong quá trình phát triển một hệ GIS. Việc thu thập và tổ chức dữ liệu, phát triển nhân sự và thiết lập các quy định cho vấn đề sử dụng GIS thường khó khăn hơn và tốn nhiều thời gian hơn.



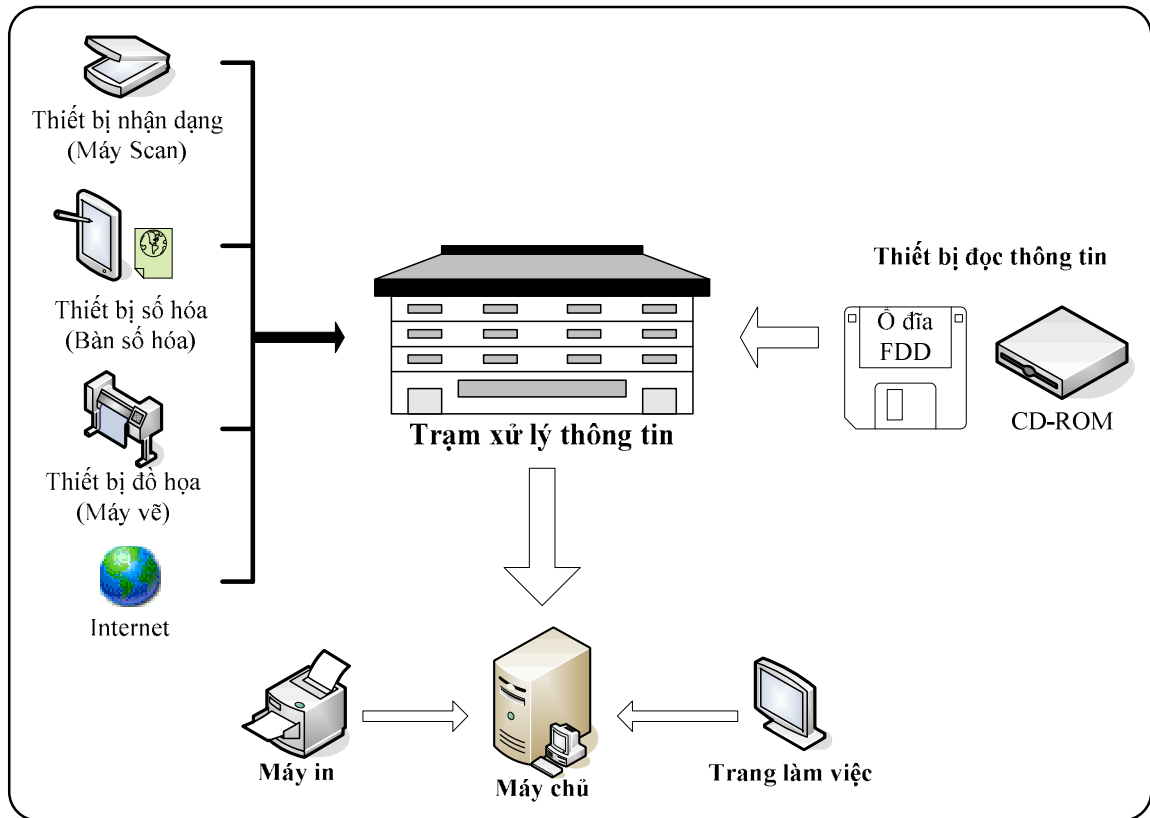


*Hình 1.6: Các thành phần của GIS*

**\* Phần cứng**

Phần cứng của GIS được xem là phần cố định mà bằng mắt thường ta có thể dễ dàng thấy được. Nó bao gồm máy tính và các thiết bị ngoại vi.

Máy tính có thể là máy có bất kỳ kích thước nào và có thể do nhiều hãng sản xuất với cấu hình khác nhau. Tuy nhiên, máy tính có cấu hình mạnh là điều mong muốn để sử dụng trong GIS. Các thiết bị ngoại vi bao gồm bàn số hóa, máy quét, máy in và máy vẽ. Các thiết bị này cũng hết sức đa dạng về kích cỡ, kiểu dáng, tốc độ xử lý và độ phân giải do các hãng khác nhau sản xuất. Chúng được kết nối với máy tính để thực hiện việc nhập và xuất dữ liệu.



**Hình 1.7:** Sơ đồ tổ chức một hệ “phần cứng GIS”

**\* Phần mềm**

Phần mềm GIS rất đa dạng và do nhiều hãng khác nhau sản xuất. Các phần mềm GIS có thể giống nhau ở chức năng, song khác nhau về tên gọi, hệ điều hành hay môi trường hoạt động, giao diện, khuôn dạng dữ liệu không gian và hệ quản trị cơ sở dữ liệu. Theo thời gian, phần mềm GIS phát triển ngày càng thân thiện với người dùng, toàn diện về chức năng và có khả năng quản lý dữ liệu hiệu quả hơn. Tuy nhiên, sự gia tăng mạnh mẽ về số lượng người bán phần mềm cũng như năng lực quản lý của GIS đã khiến cho sự lựa chọn phần mềm GIS trở thành một quyết định không đơn giản. Sự lựa chọn đó cần phải căn cứ vào mục đích sử dụng, năng lực tài chính và trình độ cán bộ. Về quy mô hay mục đích sử dụng, GIS có thể được dùng ở cấp địa phương, cấp quốc gia, khu vực hay toàn



**Hình 1.8:** Phần mềm của GIS

cầu, cho giáo dục, nghiên cứu khoa học, quy hoạch và quản lý. Do vậy, có thể chọn phần mềm tổng quát hay chuyên dụng.

Để tạo thuận lợi cho việc kết nối, chia sẻ dữ liệu, nên chọn dùng các hệ đã được tin dùng ở nhiều nơi, các hệ mở để thích ứng với những thay đổi và dễ xuất nhập, trao đổi dữ liệu với các hệ khác.

### **\* Phân dữ liệu**

Phần dữ liệu GIS bao gồm dữ liệu không gian và phi không gian. Dữ liệu không gian là dữ liệu về vị trí của các đối tượng trên mặt đất theo một hệ quy chiếu nào đó. Nó có thể được biểu diễn dưới dạng các ô lưới hay các cặp tọa độ hay cả hai, tùy thuộc vào khả năng của từng phần mềm cụ thể. Dữ liệu phi không gian là dữ liệu thuộc tính hay dữ liệu mô tả các đối tượng địa lý. Dữ liệu thuộc tính thường được trình bày dưới dạng bảng. Sự kết nối giữa dữ liệu không gian và phi không gian trong GIS là cơ sở để xác định chính xác các đối tượng địa lý và thực hiện phân tích tổng hợp GIS. Việc xây dựng một cơ sở dữ liệu GIS là một đầu tư lớn về thời gian, công sức và tiền bạc do vậy, phần dữ liệu GIS phải được quản lý khai thác một cách an toàn, tiện lợi và hiệu quả.

Với bất kỳ một hệ thông tin nào cũng phải hiểu rõ các loại dữ liệu khác nhau lưu trữ trong chúng. Dữ liệu thống kê gắn theo các hiện tượng tự nhiên với mức độ chính xác khác nhau. Hệ thống thước đo của chúng bao gồm các biên tên, số thứ tự, khoảng và tỷ lệ.

Bảng sau đây là ví dụ về các loại biến được sử dụng trong GIS.

<b>Biên tên</b> (Loại mùa màng)	<b>Thứ tự</b> (Loại cát)	<b>Khoảng</b> (Nhiệt độ)	<b>Tỷ lệ</b> (Dân số)
Lúa	Tinh	20°C	20.000
Ngô	Trung bình	25°C	40.000
Khoai tây	Thô	35°C	15.000

**Bảng 1.1:** Ví dụ các kiểu thuộc tính khác nhau trong GIS

Biên tên: là những biến chỉ có tên, không theo trật tự đặc biệt nào. Ví dụ như các loại đất sử dụng hay các loại cây trồng.

Biên thứ tự là danh sách các lớp rời rạc nhưng có trật tự. Ví dụ như trình độ học vấn tiểu học, trung học, đại học, sau đại học hay thước đo Richter của động đất, thước đo Beaufort của sức gió.

Biên khoảng cũng có trình tự tự nhiên nhưng có thêm đặc tính là khoảng cách giữa các biến còn có ý nghĩa. Ví dụ như các khoảng nhiệt độ đo.

Biên tỷ lệ, chúng có cùng đặc tính như biên khoảng nhưng chúng có giá trị 0 hay điểm bắt đầu tự nhiên. Ví dụ như lượng mưa trong mỗi tháng, thu nhập bình quân trong một năm.

Ngoài bốn loại biến dữ liệu mô tả trên, các hệ GIS còn phân chia dữ liệu thành hai lớp khác nhau là dữ liệu thuộc tính (hay dữ liệu phi không gian) và dữ liệu không gian. Hai lớp dữ liệu này được kết nối logic với nhau trong GIS. Mỗi hệ GIS đều có các công cụ lưu trữ dữ liệu thuộc tính cùng với dữ liệu không gian. Việc kết nối logic của hai loại thông tin này là rất quan trọng.

Mỗi hệ GIS cần phải hiểu được dữ liệu trong các khuôn mẫu khác nhau, không chỉ hiểu khuôn mẫu riêng của hệ thống. Ví dụ như đường biên bản đồ có thể trong khuôn mẫu tệp DXF của AutoCAD hay BNA của AtlasGIS. Thông thường, GIS hiểu ngay khuôn mẫu DXF mà không cần đến sửa đổi nào. Tương tự, GIS phải hiểu ngay khuôn mẫu DBF của các thuộc tính được lưu trữ kèm theo. Lý tưởng thì phần mềm GIS phải có khả năng đọc các khuôn mẫu dữ liệu raster như DEN, GIFF, TIFF, JPEG, EPS và khuôn mẫu vector như TIGER, HPGL, DXF, Postscript, DLG. Tuy nhiên, một số phần mềm GIS chỉ có chức năng nhập dữ liệu đơn giản.

Với dữ liệu ba chiều, phần lớn các phần mềm GIS trợ giúp lưới tam giác không đều (TIN). Một số khác trợ giúp cấu trúc raster trên cơ sở lưới bao gồm cây tứ phân và khả năng chuyển đổi toàn bộ hay một phần dữ liệu vào cấu trúc này. Một vài phần mềm GIS chỉ trợ giúp khuôn mẫu riêng, chúng phụ thuộc vào nhà sản xuất phần mềm. Những năm gần đây, các thao tác hỗ trợ của GIS được xây dựng để hệ thống hiểu được dữ liệu trong khuôn mẫu chuẩn. Một số chuẩn chuyển đổi được xây dựng theo mức quốc gia hay quốc tế như SDTS (Spatial Data Transfer Standard) hay DIGEST.

### **\* *Phần chuyên gia***

Trong GIS, phần con người còn được biết đến dưới các tên gọi khác như phần não hay phần sống của hệ thống. Con người tham gia vào việc thiết lập, khai thác và bảo trì hệ thống một cách gián tiếp hay trực tiếp. Có hai nhóm người quan trọng trực tiếp quyết định sự tồn tại và phát triển của GIS là người sử dụng và người quản lý sử dụng GIS.

Đội ngũ những người sử dụng GIS bao gồm các thao tác viên, kỹ thuật viên hỗ trợ kỹ thuật và các chuyên gia về các lĩnh vực khác nhau có sử dụng thông tin địa lý. Người sử dụng trở thành một thành phần của GIS khi tiến hành những phép phân tích phức tạp, các thao tác phân tích không gian và mô hình hóa. Công việc này yêu cầu các kỹ năng để chọn lựa và sử dụng các công cụ từ hộp công cụ của GIS và có kiến thức về các dữ liệu đang được sử dụng. Hiện tại và trong những năm trước mắt, GIS vẫn sẽ phụ thuộc vào người sử dụng có nắm vững kiến thức về những gì họ đang làm chứ không đơn giản chỉ ấn một nút là đủ.

✓ ***Người sử dụng hệ thống:*** là những người sử dụng GIS để giải quyết các vấn đề không gian. Họ thường là những người được đào tạo tốt về lĩnh vực GIS hay GIS chuyên dụng. Nhiệm vụ chủ yếu của họ là số hóa bản đồ, kiểm tra lỗi, soạn thảo, phân tích dữ liệu thô và đưa ra các giải pháp cuối cùng để truy vấn dữ liệu địa lý. Dù được đào tạo chính qui hay tại chức thì người sử dụng hệ thống vẫn



phải được thường xuyên đào tạo lại vì phần mềm GIS thay đổi liên tục và do yêu cầu mới của kỹ thuật phân tích.

✓ **Thao tác viên hệ thống:** có trách nhiệm vận hành hệ thống hàng ngày để người sử dụng hệ thống làm việc hiệu quả. Công việc của họ là sửa chữa khi chương trình bị tắt nghẽn hay là công việc trợ giúp nhân viên thực hiện các phân tích có độ phức tạp cao. Đôi khi họ còn có trách nhiệm huấn luyện người dùng, họ cũng là người có kinh nghiệm như người sử dụng hệ thống. Họ hiểu biết về cấu hình phần mềm và phần cứng để có thể yêu cầu nâng cấp. Họ còn làm việc như người quản trị hệ thống, quản trị cơ sở dữ liệu, an toàn, toàn vẹn cơ sở dữ liệu để tránh hư hỏng, mất mát dữ liệu.

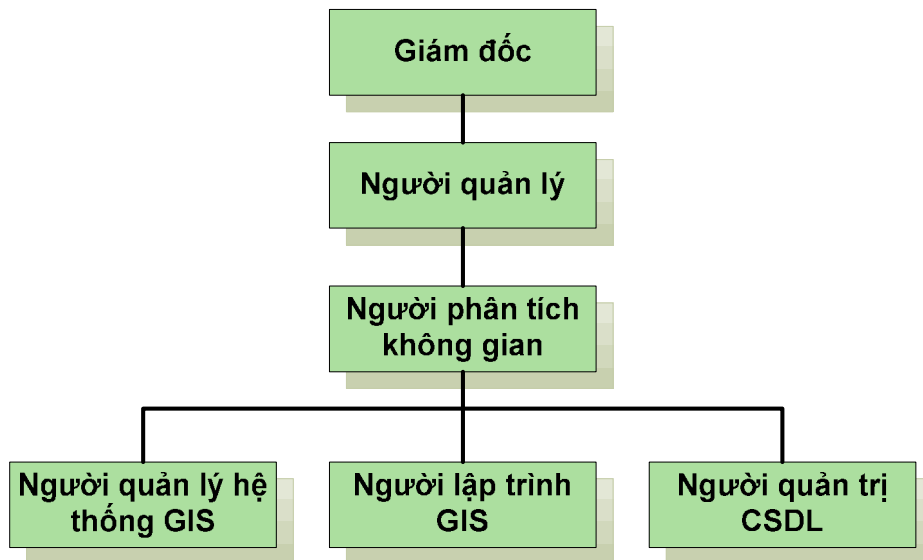
✓ **Nhà cung cấp GIS:** có trách nhiệm cung cấp phần mềm, cập nhật phần mềm, phương pháp nâng cấp cho hệ thống. Đôi khi tham gia huấn luyện người dùng GIS thông qua các hợp đồng với quản trị hệ thống.

✓ **Nhà cung cấp dữ liệu:** có thể là tổ chức Nhà nước hay tư nhân. Thông thường, các công ty tư nhân cung cấp dữ liệu sửa đổi từ dữ liệu các cơ quan Nhà nước để cho phù hợp với ứng dụng cụ thể. Thường thì các cơ quan Nhà nước cung cấp dữ liệu được xây dựng cho chính nhu cầu của họ, nhưng dữ liệu này có thể được sử dụng trong các tổ chức, cơ quan khác. Một số dữ liệu này được bán với giá rẻ hay cho không đổi với các dự án GIS phi lợi nhuận.

✓ **Người phát triển ứng dụng:** là những người lập trình viên được đào tạo. Họ xây dựng các giao diện người dùng, làm giảm khó khăn khi thực hiện các thao tác cụ thể trên các hệ thống GIS chuyên nghiệp. Phần lớn, lập trình GIS bằng ngôn ngữ macro do nhà cung cấp GIS xây dựng để người phát triển ứng dụng có khả năng ghép nối với các ngôn ngữ máy tính truyền thống.

✓ **Chuyên viên phân tích hệ thống GIS:** là nhóm người chuyên nghiên cứu thiết kế hệ thống. Phần lớn họ là đội ngũ chuyên nghiệp, có trách nhiệm xác định mục tiêu của hệ GIS trong cơ quan, hiệu chỉnh hệ thống, đề xuất kỹ thuật, phân tích đúng đắn, đảm bảo tích hợp tốt hệ thống trong cơ quan. Thông thường, chuyên gia phân tích hệ thống là nhân viên của các hãng lớn chuyên về cài đặt GIS.

Tóm lại, một dự án GIS chỉ thành công khi nó được quản lý tốt và con người tại mỗi công đoạn phải có kỹ năng tốt. Dưới đây là một ví dụ về cấu trúc quản lý dự án GIS độc lập. Tuy nhiên, trong nhiều trường hợp có thể kết hợp quản lý dự án GIS với cấu trúc quản lý có sẵn trong cơ quan.



*Hình 1.9: Sơ đồ quản lý dự án GIS*

### **1.2.2. Chức năng của GIS**

Với thành phần như trên, GIS có thể và phải đảm đương các chức năng chủ yếu sau:

#### **✓ Nhập dữ liệu**

Nhập dữ liệu là một chức năng của GIS qua đó dữ liệu dưới dạng tương tự hay dạng số được biến đổi sang dạng số có thể sử dụng được bằng GIS. Việc nhập dữ liệu được thực hiện nhờ vào các thiết bị như bàn số hóa, máy quét, bàn phím và các chương trình hay môđun nhập và chuyển đổi dữ liệu của GIS.

#### **✓ Quản lý dữ liệu**

Việc xây dựng một cơ sở dữ liệu GIS lớn bằng các phương pháp nhập dữ liệu khác nhau thường rất tốn kém về thời gian, công sức và tiền bạc. Số chi phí bằng tiền cho việc xây dựng cơ sở dữ liệu có thể lớn hơn hẳn chi phí phần cứng và phần mềm GIS. Điều đó phần nào nói lên ý nghĩa của việc quản lý dữ liệu, một chức năng quan trọng của tất cả các hệ thống thông tin địa lý. Chức năng này bao gồm việc tổ chức lưu trữ và truy cập dữ liệu sao cho hiệu quả nhất.

#### **✓ Phân tích dữ liệu**

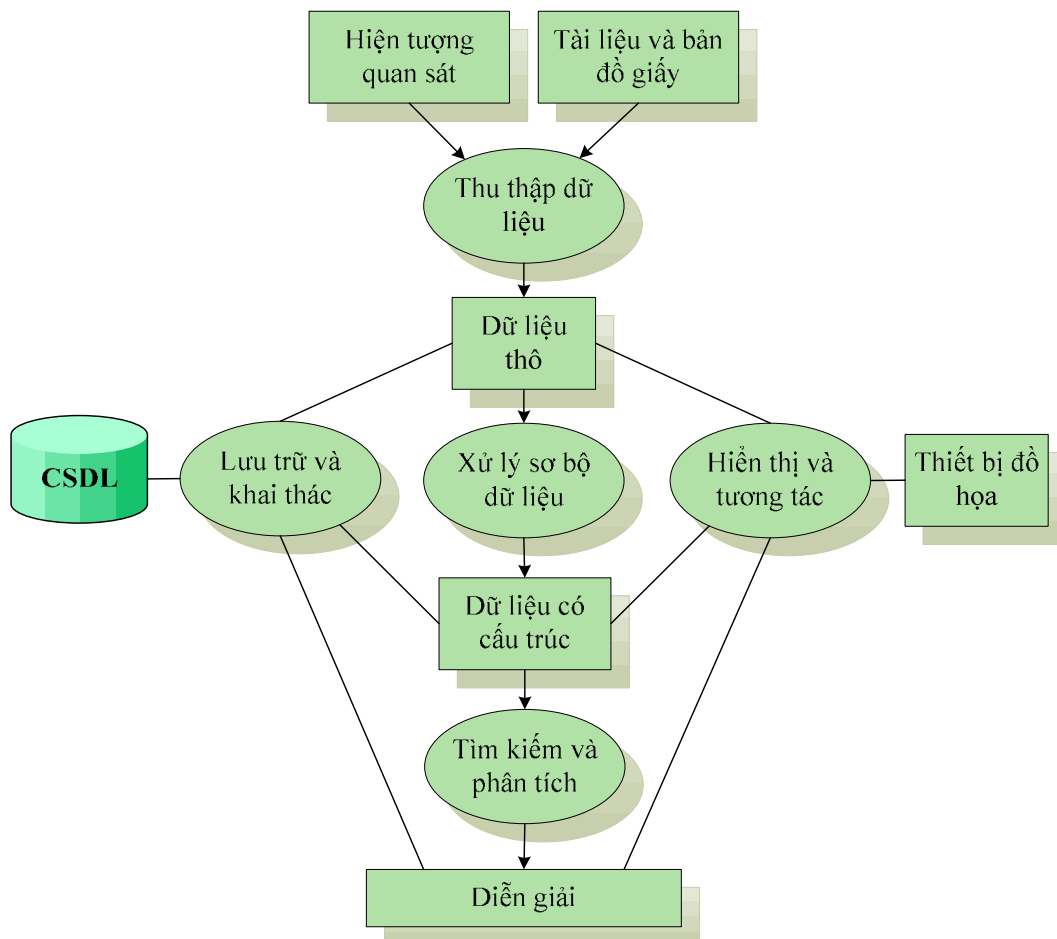
Phân tích dữ liệu là chức năng quan trọng nhất của GIS. GIS cung cấp các công cụ cần thiết để phân tích dữ liệu không gian, dữ liệu thuộc tính và phân tích tổng hợp cả hai loại dữ liệu đó ở trong cơ sở dữ liệu để tạo ra thông tin mới trợ giúp các quyết định mang tính không gian.

#### **✓ Xuất dữ liệu**

Chức năng xuất dữ liệu hay còn gọi là chức năng báo cáo của GIS cho phép hiển thị, trình bày các kết quả phân tích và mô hình hóa không gian bằng GIS dưới

dạng bản đồ, bảng thuộc tính hay văn bản trên màn hình hay trên các vật liệu truyền thống khác ở các tỷ lệ và chất lượng khác nhau tùy thuộc vào yêu cầu của người dùng và khả năng của các thiết bị xuất dữ liệu như màn hình, máy in và máy vẽ.

Sức mạnh của các chức năng của hệ thống GIS khác nhau là khác nhau. Kỹ thuật xây dựng các chức năng cũng rất khác nhau. Sơ đồ sau đây mô tả quan hệ giữa các nhóm chức năng và cách biểu diễn thông tin khác nhau của GIS.



**Hình 1.10:** Các nhóm chức năng trong GIS

Chức năng thu thập dữ liệu tạo ra dữ liệu từ các quan sát hiện tượng thế giới thực và từ các tài liệu, bản đồ giấy, đôi khi chúng có sẵn dưới dạng số. Kết quả ta có tập dữ liệu thô, có nghĩa là dữ liệu này không được phép áp dụng trực tiếp cho chức năng truy nhập và phân tích của hệ thống. Chức năng xử lý sơ bộ dữ liệu sẽ biến đổi dữ liệu thô thành dữ liệu có cấu trúc để sử dụng trực tiếp các chức năng tìm kiếm và phân tích không gian. Kết quả tìm kiếm và phân tích được xem như diễn giải dữ liệu, đó là tổ hợp hay biến đổi đặc biệt của dữ liệu có cấu trúc. Hệ thống thông tin địa lý phải có phần mềm công cụ để tổ chức và lưu trữ các loại dữ liệu khác nhau, từ dữ liệu thô đến dữ liệu diễn giải. Phần mềm công cụ này phải có các thao tác lưu trữ, truy nhập; đồng thời có khả năng hiển thị, tương tác đồ họa với tất cả các loại dữ liệu.

### 1.3. Một số ứng dụng GIS

Kể từ khi ra đời cho đến nay, GIS đã được ứng dụng ở nhiều nơi trên thế giới, trong nhiều lĩnh vực và ở các quy mô khác nhau. Các ứng dụng đầu tiên của GIS ở các nước trên thế giới không giống nhau.

Ở Châu Âu, xu hướng chủ yếu là ứng dụng GIS vào việc xây dựng các hệ thống quản lý đất đai và cơ sở dữ liệu cho môi trường.

Ở Canada, nơi chứng kiến sự ra đời của GIS cấp quốc gia đầu tiên trên thế giới, một ứng dụng trong lâm nghiệp quan trọng của GIS là xây dựng kế hoạch khai thác gỗ, xác định các con đường để đi khai thác gỗ và báo cáo kết quả cho chính phủ địa phương.

Ở Mỹ, GIS được ứng dụng trong rất nhiều lĩnh vực. Một dự án đang được đề cập đến về việc sử dụng công nghệ GIS là TIGER (Topographically Integrated Geographical Referencing) do cơ quan điều tra dân số và sở địa chất Mỹ triển khai. Dự án này được thiết kế để tạo thuận lợi cho cuộc điều tra dân số năm 1990 và đã được phát triển để xây dựng được mô hình máy tính hóa cho mạng lưới giao thông Mỹ với trị giá khoảng 170 triệu đôla.

Ở Trung Quốc và Nhật Bản, GIS được ứng dụng chủ yếu vào việc xây dựng mô hình và quản lý các thay đổi của môi trường do mức độ nghiêm trọng của thiên tai.

Ở các nước đó, các lĩnh vực ứng dụng của GIS hết sức đa dạng và ngày càng gia tăng cùng với sự phát triển của công nghệ và sự xuất hiện các vấn đề mới ở các quy mô khác nhau. GIS đã được áp dụng vào lập bản đồ các vùng sinh thái nông nghiệp, lập bản đồ thích hợp đất đai, dự báo sản lượng, quy hoạch và quản lý sử dụng đất.

Trong lâm nghiệp, GIS đã được sử dụng để nhập, lưu trữ, quản lý và phân tích các bản đồ rừng để phục vụ việc khai thác, bảo vệ và phát triển rừng.

Trong lĩnh vực khảo cổ học, các kỹ thuật GIS được sử dụng để phân tích các địa điểm đã biết và dự báo vị trí các điểm khảo cổ chưa được phát hiện.

Với khả năng liên kết các lớp dữ liệu khác nhau, GIS được sử dụng có hiệu quả trong việc tìm kiếm khoáng sản trên cơ sở tổng hợp các dữ liệu viễn thám, địa vật lý, địa hóa và địa chất.

Ở các đô thị, GIS đã được sử dụng để trợ giúp các quyết định pháp lý, hành chính, kinh tế cũng như các hoạt động quy hoạch khác.

Bên cạnh các ứng dụng ở quy mô địa phương, quốc gia, GIS cũng đã được ứng dụng ở quy mô liên quốc gia và toàn cầu. Một ví dụ điển hình là hệ ARC/INFO của ESRI đã được chọn dùng trong chương trình CORINE (Coordinated Information on the European Environment) do Cộng đồng Châu Âu

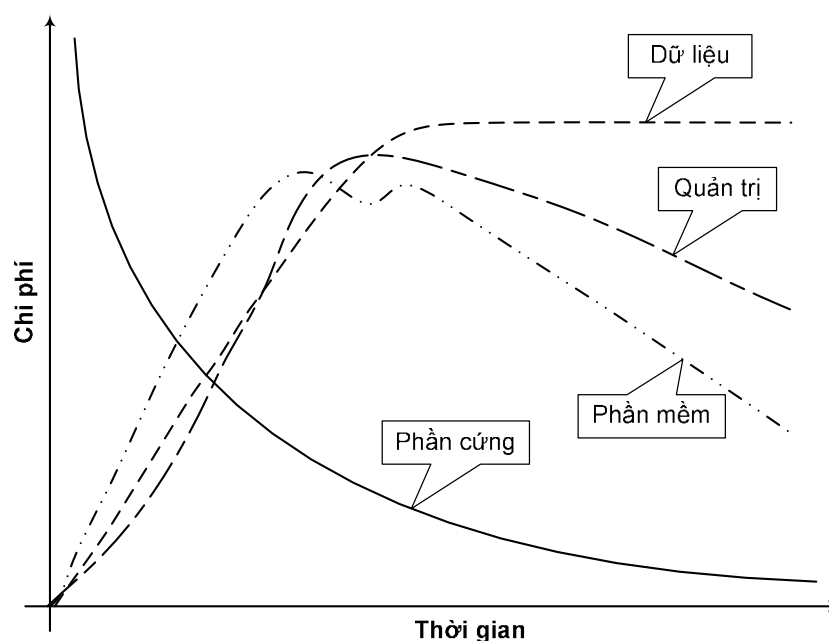
khởi xướng năm 1985. Hệ thống đã hoạt động thành công cho phép người sử dụng ở các quốc gia khác nhau tiếp cận hệ thống và trao đổi dữ liệu. Các bộ dữ liệu đất, khí hậu, địa hình và sinh thái đã được phát triển và các dự án được xúc tiến để phân tích các vấn đề môi trường cụ thể liên quan đến khí thải, ô nhiễm nước và xói mòn đất.

Một ví dụ khác là vào năm 1983, chương trình môi trường liên hợp quốc (UNEP) đã chọn ESRI để xây dựng một hệ thống dựa vào GIS để phân tích và lập bản đồ các vùng sa mạc trên quy mô toàn cầu. Tiếp đó, năm 1985, UNEP đã xúc tiến việc xây dựng cơ sở dữ liệu tài nguyên toàn cầu (GRID) với sự hỗ trợ của GIS.

Trên đây là một số ví dụ điển hình về ứng dụng GIS, chắc chắn với thời gian và sự phát triển của công nghệ, các ứng dụng của GIS sẽ ngày càng phong phú hơn, bao gồm tất cả các ứng dụng truyền thống cũng như các ứng dụng mới vì sự phát triển chung của nhân loại.

## 1.4. Chi phí cho GIS

GIS cần phần cứng, phần mềm, dữ liệu và quản trị để dẫn đến thành công của mọi ứng dụng. Tuy nhiên, mỗi loại đều đòi hỏi nguồn kinh phí nhất định. Hiện nay, rất nhiều phần mềm GIS được thương mại hóa cho nên giá thành của chúng hạ nhanh chóng nhưng chức năng của chúng lại ngày càng được tăng cường. Chỉ có chi phí cho dữ liệu không gian hầu như vẫn giữ nguyên ở mức cao. Thông thường, khoảng 70% tổng chi phí của dự án GIS là dành cho việc thu thập dữ liệu.



**Hình 1.11:** Chi phí cho GIS

## CHƯƠNG 2

---

# Các mô hình dữ liệu

### 2.1. Giới thiệu chung về dữ liệu

Trong phần này, chúng ta sẽ cùng hệ thống hóa lại một số khái niệm về dữ liệu để có cơ sở tìm hiểu tiếp các vấn đề về tổ chức dữ liệu và cơ sở dữ liệu GIS.

Vậy dữ liệu là gì? Dữ liệu là một khái niệm rất quan trọng trong các vấn đề lưu trữ và xử lý thông tin. Trong một nghĩa giới hạn thì dữ liệu là sự thể hiện của thông tin, thể hiện này phải ở dạng thích hợp cho việc thao tác, truyền phát bởi các phương tiện tự động, thường là các hệ máy tính. Trong các hệ máy tính thì dữ liệu được lưu trữ ở dạng nhị phân. Trong việc xử lý thông tin trên máy tính, thuật ngữ “*data*” thường để chỉ các thông tin được xử lý chứ không chỉ các phần mềm, các chương trình xử lý những dữ liệu này. Tuy nhiên, nhiều khi thuật ngữ “*data*” cũng được dùng với nghĩa rộng hơn để chỉ mọi dạng thông tin bao gồm cả các chương trình.

Theo Homby (1988), dữ liệu là:

- Các sự thật, các thứ biết chắc chắn và từ đó có thể rút ra các kết luận.
- Thông tin được chuẩn bị và được thao tác trên chương trình máy tính.

Ngân hàng dữ liệu là trung tâm có hồ sơ toàn diện về các dữ liệu máy tính. Cơ sở dữ liệu là kho dữ liệu đã được máy tính hóa. Xử lý dữ liệu là thực hiện các thao tác trên dữ liệu để thu được thông tin, lời giải cho các vấn đề...

Trong GIS, một hệ thống dựa vào máy tính, ta có thể hiểu dữ liệu là sự phản ánh thế giới thực theo một ngôn ngữ hay một quy ước nào đó thông qua những quan sát, đo đạc và tính toán được ghi lại trong thời gian nghiên cứu. Dữ liệu trong GIS là dữ liệu địa lý và thường được đề cập đến dưới dạng các cặp dữ liệu sau:

- Dữ liệu tương tự và dữ liệu số
- Dữ liệu không gian và dữ liệu phi không gian
- Dữ liệu tự nhiên và dữ liệu kinh tế - xã hội

Dữ liệu tương tự là các dữ liệu được thể hiện hay lưu trữ trên các vật liệu truyền thống như giấy ảnh, bản đồ, báo cáo... mà bằng mắt thường hay các loại kính phóng đại ta có thể thấy được.

Dữ liệu số là dữ liệu được mã hóa bằng số theo hệ nhị phân và được lưu trữ trong máy tính, máy ảnh số, trong băng từ, đĩa cứng, đĩa mềm, đĩa quang...

Dữ liệu không gian là dữ liệu về vị trí của người, vật, hiện tượng hay quá trình theo một hệ tọa độ hay hệ quy chiếu nào đó. Dữ liệu này cho ta biết các đối tượng nghiên cứu ở đâu.

Dữ liệu phi không gian là dữ liệu thuộc tính hay mô tả gắn với các vị trí của các đối tượng địa lý khác nhau. Dữ liệu này có thể là định tính hay định lượng và nó cho ta biết các đối tượng nghiên cứu là cái gì.

Dữ liệu tự nhiên là dữ liệu về các đối tượng tự nhiên như đất, đá, sông ngòi, khí hậu... còn dữ liệu kinh tế - xã hội là dữ liệu về sự tồn tại và phát triển của con người như dân số, việc làm, y tế, trường học, giao thông...

## **2.2. Giới thiệu chung về bản đồ**

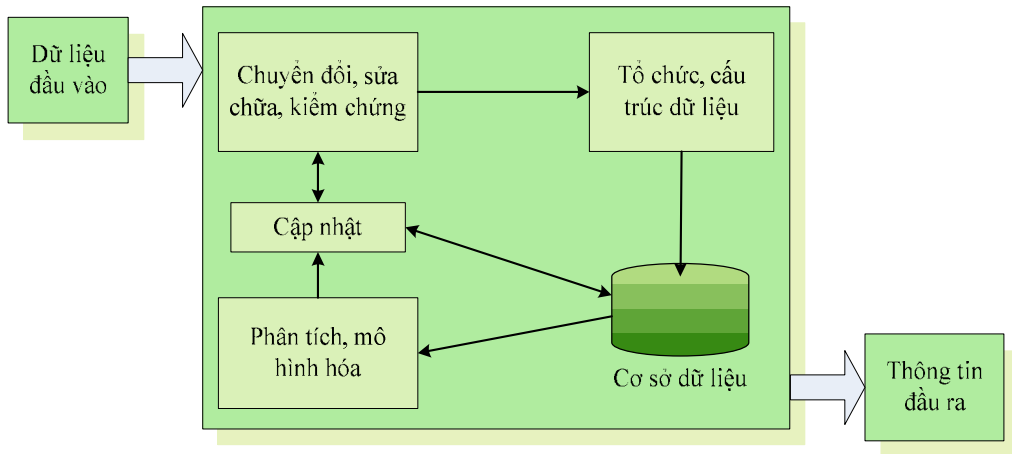
### ***2.2.1. Khái niệm bản đồ***

Đã từ lâu con người đã biết lập ra và sử dụng bản đồ phục vụ cho các hoạt động dân sự và quân sự của mình. Bản đồ được xem như là một công cụ truyền thông để biểu diễn hay mô hình hóa các sự vật, hiện tượng hay quá trình trên bề mặt Trái đất ở các tỷ lệ và hệ quy chiếu khác nhau. Bản đồ là một ví dụ điển hình của loại dữ liệu tương tự. Bản đồ thường chứa hai dạng thông tin cơ bản sau:

- Thông tin không gian mô tả vị trí và hình dạng của các đối tượng địa lý và mối liên hệ không gian giữa chúng.
- Thông tin mô tả về các đối tượng

Bản đồ là nguồn dữ liệu, thông tin địa lý quan trọng, là đầu vào và đầu ra, là nguyên vật liệu và là sản phẩm của GIS. Với GIS, chúng ta có một công cụ mới để nhập, lưu trữ, cập nhật, khai thác, sử dụng và lập bản đồ tiện lợi, hiệu quả hơn.

Khi xây dựng GIS, việc đầu tiên là phải tạo ra một cơ sở dữ liệu bản đồ số. Để tự động hóa các bản đồ, chúng ta phải nghĩ đến việc lưu trữ các thông tin một cách rõ ràng và rành mạch, cấu trúc dữ liệu ra sao và sử dụng cơ sở dữ liệu bản đồ như thế nào. GIS có một mô hình dữ liệu rõ ràng cho việc biểu diễn các bản đồ trong máy tính. Một khi chúng ta đã hiểu được các bản đồ được tạo và lưu trữ như thế nào, chúng ta có thể bắt tay vào việc xây dựng cơ sở dữ liệu cần thiết cho việc tiến hành thực hiện một hệ GIS.



**Hình 2.1:** Sơ đồ chuyển hóa dữ liệu thành thông tin

### 2.2.2. Các đối tượng bản đồ

Một bản đồ sẽ hiển thị thông tin bằng cách biểu diễn đồ họa các thành phần của nó. Các thông tin vị trí sẽ được biểu diễn bằng các điểm cho các đối tượng như trạm điện thoại, đài truyền hình; các đường cho các đối tượng như đường phố, sông ngòi và các vùng cho các đối tượng như hồ, các khu bảo tồn.

#### ✓ **Đối tượng điểm**

Đối tượng điểm là biểu diễn đơn giản nhất về một đối tượng, được dùng để biểu diễn cho một vị trí riêng biệt xác định một đối tượng bản đồ mà hình dạng của nó quá nhỏ để có thể biểu diễn như một đối tượng đường hay vùng hoặc nó có thể biểu diễn một địa điểm không có diện tích. Một ký hiệu đặc biệt hay một nhãn thường mô tả một điểm



#### ✓ **Đối tượng đường**

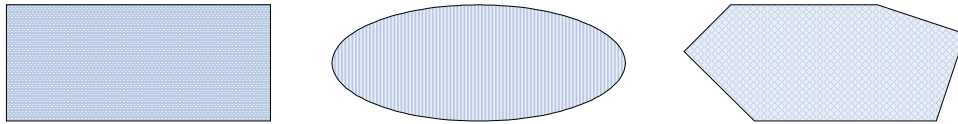
Một đối tượng đường là một tập hợp có thứ tự các tọa độ mà khi được liên kết với nhau sẽ biểu diễn một hình tuyến tính của một đối tượng bản đồ quá hẹp để có thể hiển thị như một vùng hoặc có thể coi nó như một đối tượng không có độ rộng.



#### ✓ **Đối tượng vùng**

Một đối tượng vùng là một hình khép kín mà biên của nó bao quanh một vùng đồng nhất như một quốc gia, một tỉnh...





### ✓ *Các quan hệ không gian*

Các mối quan hệ không gian giữa các đối tượng được thể hiện trên bản đồ phụ thuộc vào khả năng của người xem bản đồ. Ví dụ bạn có thể nhìn vào một bản đồ và biết được có một thành phố ở cạnh một hồ nào đó hay không, biết được khoảng cách tương đối giữa các thành phố theo các con đường cũng như xác định được đường đi ngắn nhất giữa hai thành phố, nhận biết được một bệnh viện ở gần nhất và con đường đi tới đó... Loại thông tin này không được thể hiện chính xác trên bản đồ.

### ✓ *Các ký hiệu và các nhãn thể hiện thông tin mô tả*

Cũng như các hình ảnh đồ họa, bản đồ thể hiện vị trí của các đối tượng cùng với các đặc tính để dễ hiểu hơn. Các đặc tính của đối tượng bản đồ được thể hiện bằng các ký hiệu. Ví dụ các con đường khác nhau được vẽ bằng các màu sắc, độ rộng, mẫu khác nhau, các con suối được vẽ bằng đường màu xanh và được gắn thêm tên của chúng, và các trường học được thể hiện bằng ký hiệu, các hồ được tô bằng màu xanh... Bằng cách này, các đối tượng được thể hiện tương tự như thông tin mô tả của chúng.

## **2.2.3. Các hệ quy chiếu và hệ tọa độ sử dụng ở Việt Nam**

Trước năm 2000, theo quy định của Tổng Cục Địa chính, bản đồ Việt Nam sử dụng hệ quy chiếu và hệ tọa độ Hà Nội-72 với lưới chiếu Gauss là một lưới chiếu hình trụ ngang giữ góc. Theo phép chiếu Gauss thì bề mặt ellipsoid được chia thành 60 múi, mỗi múi 6 độ. Sau khi căng ra mặt phẳng thì mỗi múi này từ một nhị giác cầu trở thành một nhị giác phẳng giới hạn bởi hai cung kinh tuyến biên.

Đến năm 2000, Thủ tướng Chính phủ đã ra quyết định số 83/2000QĐ-TTg ngày 12/7/2000 về việc sử dụng hệ quy chiếu và hệ tọa độ quốc gia mới thống nhất trên phạm vi toàn quốc cho tất cả các loại tư liệu đo đạc bản đồ, thay thế cho hệ quy chiếu và hệ tọa độ quốc gia Hà Nội-72 được sử dụng trước đây. Trong quyết định có nêu rõ các nội dung chính liên quan đến hệ quy chiếu và hệ tọa độ quốc gia mới như sau:

- Tên hệ quy chiếu và hệ tọa độ quốc gia mới là VN-2000.
- Lựa chọn hệ quy chiếu quốc tế WGS-84 toàn cầu làm ellipsoid quy chiếu quốc gia.
- Lựa chọn điểm gốc tọa độ quốc gia là điểm N00 đặt tại khuôn viên Viện nghiên cứu Địa chính (Viện Khoa học Công nghệ Địa chính trước đây), đường Hoàng Quốc Việt, Hà Nội.

- Lựa chọn lưới chiếu tọa độ phẳng UTM (Universal Transverse Meleator) quốc tế là lưới chiếu tọa độ phẳng quốc gia. UTM cũng là lưới chiếu hình trụ ngang giữ góc với các múi 6 độ như lưới chiếu Gauss. Tuy nhiên, điểm khác nhau giữa hai lưới chiếu này là trong phép chiếu hình của Gauss, mặt trụ tiếp xúc với bề mặt của ellipsoid theo đường kinh tuyến giữa của mỗi múi, còn ở lưới chiếu UTM, mặt trụ cắt bề mặt của ellipsoid theo hai vòng đứng cách kinh tuyến giữa 2 độ.
- Việc chia múi và phân mảnh hệ thống bản đồ cơ bản được thực hiện theo hệ thống lưới chiếu hình trụ ngang đồng góc UTM quốc tế.

## 2.3. Cấu trúc dữ liệu

Cấu trúc dữ liệu đề cập đến cách thức tổ chức dữ liệu thành các file dữ liệu. Cho đến nay, trong GIS, người ta thường nói đến hai loại cấu trúc dữ liệu không gian là vector và raster. Điều đó có nghĩa là các dữ liệu không gian có thể được mã hóa, được lưu trữ trong máy tính theo hai cấu trúc kể trên.

### 2.3.1. Cấu trúc dữ liệu vector

Như ta biết, vector là một đại lượng biến thiên có độ lớn, có hướng và có thể phân tích thành các hợp phần. GIS vector là hệ thống sử dụng biểu diễn vector trong lưu trữ và phân tích dữ liệu.

Với mô hình vector, toàn bộ thế giới thực hay các đối tượng địa lý đều có thể được biểu diễn được bằng ba loại thực thể không gian cơ sở sau:

- Điểm
- Đường
- Vùng hay đa giác

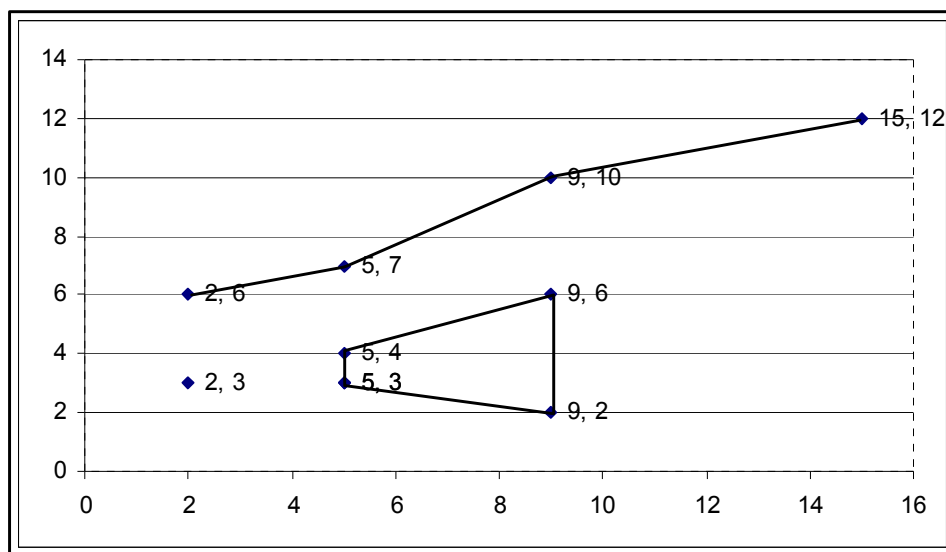
Các thực thể đó được mô tả hình học bằng cách ghi lại các cặp tọa độ  $x, y$  và có thể cả  $z$  (đối với GIS 3 chiều) theo một hệ quy chiếu nhất định (hệ tọa độ phẳng hay hệ tọa độ địa lý).

Một điểm được biểu diễn bằng một cặp tọa độ duy nhất;  $P = (x, y)$ .

Một đường được biểu diễn bằng một danh sách các cặp tọa độ nối tiếp nhau;  $L = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) = P_1, P_2, \dots, P_n$ .

Một đa giác được biểu diễn bằng một danh sách các cặp tọa độ nối tiếp nhau và khép kín hay danh sách các đường nối tiếp nhau và khép kín;  $P = L_1, L_2, \dots, L_n$ .

Các đối tượng trên bề mặt Trái đất được thể hiện trên bản đồ theo một mặt phẳng, bản đồ hai chiều như điểm, đường, vùng. Hệ tọa độ x,y (Cartesian) dùng để qui chiếu các vùng bản đồ tương ứng với các vùng trên mặt đất.

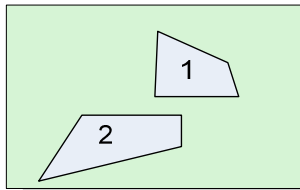
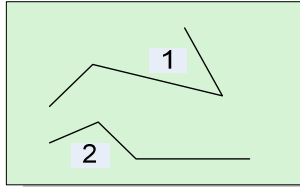
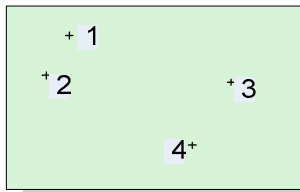


**Hình 2.2:** Hệ tọa độ phẳng

Theo hệ tọa độ này, mỗi điểm được lưu trữ bằng một cặp tọa độ (x,y). Các đường được lưu trữ bằng một dãy các cặp tọa độ (x,y). Các vùng được ghi thành một dãy các cặp tọa độ (x,y) xác định các đoạn thẳng bao quanh vùng đó. Với các cặp tọa độ (x,y), ta có thể biểu diễn các điểm, đường, vùng như là một dãy các tọa độ thay cho các hình ảnh hoặc đồ thị. Trong hình vẽ trên, cặp tọa độ (2,3) biểu diễn vị trí một điểm; các cặp tọa độ (2,6) (5,7) (9,10) (15,12) biểu diễn một cung; các cặp tọa độ (5,3) (5,4) (9,6) (9,2) (5,3) biểu diễn cho một vùng. Chú ý là cặp tọa độ đầu tiên và cuối cùng phải trùng nhau vì một vùng luôn có hình khép kín. Các tọa độ biểu diễn cho các đối tượng bản đồ sẽ được lưu trữ như một tập các số x,y trong máy tính và chính vì vậy mà xuất hiện thuật ngữ số hóa bản đồ.

Các tọa độ có thể được biểu diễn bằng các đơn vị như inch, cm, m... để đo các khoảng cách trên bản đồ. Nhưng các bản đồ lại hay sử dụng các hệ tọa độ thế giới thực được qui chiếu trên một mặt phẳng. Các tọa độ này biểu diễn một vị trí thực sự trên bề mặt Trái đất trong một vài hệ tọa độ. Trong ví dụ trên, các tọa độ thực được qui chiếu vào trong một hệ tọa độ gọi là UTM (Universal Transverse Meleator) trong đó đơn vị tọa độ là mét.

Việc lưu trữ các tọa độ của một đối tượng có thể được thực hiện như trên. Tuy nhiên, khi ta có nhiều đối tượng, ta có thể gán cho mỗi đối tượng một số nguyên liên tiếp nhau hoặc một định danh riêng biệt (ID). Khi đó các tọa độ sẽ được lưu trữ một cách tương ứng với định danh đó. Ví dụ:



Điểm số	Toạ độ x,v
1	3,8
2	2,7
3	6,7
4	4,5

Đường số	Toạ độ x,v
1	1.5 3.6 6.5 6.7
2	1.1 3.3 5.2 8.2

Vùng số	Toạ độ x,y
1	3,6 5,7 7,6 4,4 3,6
2	7,3 8,5 9,3 1,8 7,3

Như vậy, các thực thể không gian trong mô hình vector ít nhiều sát với các thực thể không gian trên Trái đất

**\* Ưu điểm của cấu trúc dữ liệu vector:**

- Tiết kiệm bộ nhớ.
- Dễ biểu diễn các quan hệ không gian.
- Thích hợp với phân tích mạng.
- Dễ tạo đồ họa đẹp, chính xác.

**\* Nhược điểm của cấu trúc dữ liệu vector:**

- Cấu trúc phức tạp.
- Khó chồng ghép.
- Khó biểu diễn không gian liên tục.

### 2.3.2. Cấu trúc dữ liệu raster

Từ “raster” bắt nguồn từ tiếng Đức và thường được dùng như một thuật ngữ để mô tả một hệ thống biểu diễn ảnh trong đó ảnh bao gồm các ô nhỏ, đồng nhất ở bên trong được sắp xếp thành một lưới.

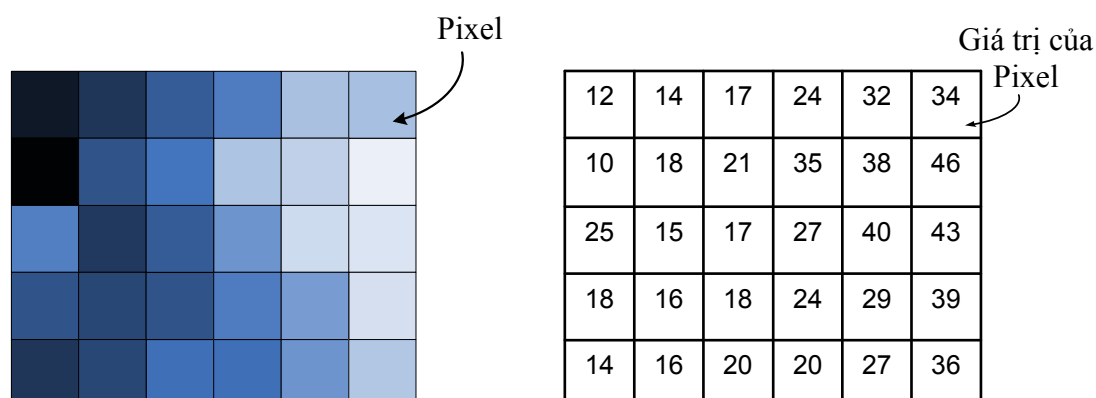
Ta có thể tạo ra một mô hình dữ liệu raster bằng cách chồng một lớp đều hay không đều lên trên các đối tượng địa lý hay nói cách khác ta có thể biểu diễn thế giới thực bằng các lưới có độ phân giải khác nhau.

Có nhiều lưới đều như lưới lục giác, tam giác và ô vuông trong đó lưới ô vuông được dùng phổ biến nhất. Trong GIS, mô hình raster thường được dùng để chỉ một lưới các ô vuông đều có các giá trị thuộc tính độc lập.

Khác với mô hình vector, mô hình dữ liệu raster có các đặc điểm sau:

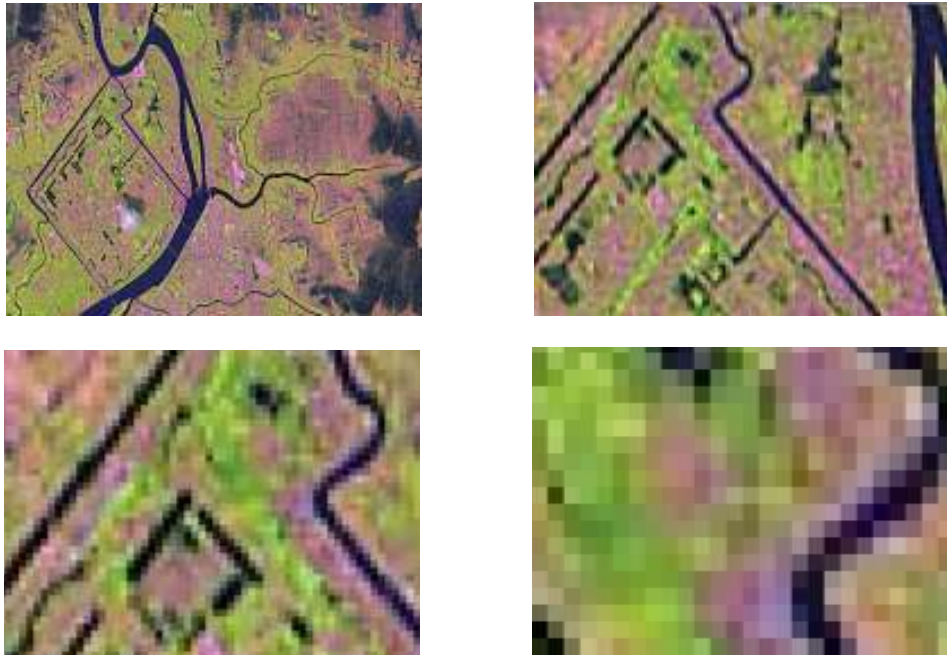
- Không gian được chia thành các ô.
- Vị trí của các đối tượng địa lý được xác định bởi vị trí dòng và cột của các ô mà chúng chiếm đóng.
- Độ phân giải không gian được quyết định bởi kích thước ô.
- Ô hay pixel là đơn vị cơ sở trong mô hình raster.

Mô hình dữ liệu raster giống một bức ảnh hơn là một bản đồ. Nếu nhìn vào một bức ảnh qua kính phóng đại, ta sẽ thấy nó được tạo bởi một loạt các chấm điểm có màu hay xám độ khác nhau. Mô hình dữ liệu raster cũng như vậy; nó là một lưới đều đặn các chấm điểm (gọi là các ô hay pixel) được điền bằng các giá trị. Dùng mô hình dữ liệu raster, Trái đất được xem như một bề mặt liên tục.



Có ba cách để giải đoán một chấm điểm trong ảnh. Cách thứ nhất là phân loại mỗi chấm điểm như là thuộc cái gì đó; một nhóm các pixel được phân loại tương tự trở thành một đối tượng. Cách thứ hai là chỉ việc đo giá trị màu hay xám độ của nó. Cách thứ ba là xác định pixel tương đối so với một điểm quy chiếu, như mực nước biển trung bình đối với độ cao.

Trong mô hình dữ liệu raster, mỗi vị trí được biểu diễn như một ô. Ma trận các ô được tổ chức thành dòng và cột, được gọi là lưới (Grid). Các giá trị ô là các số biểu diễn dữ liệu danh định như các lớp sử dụng đất, các số đo cường độ ánh sáng hay các số đo tương đối.



**Hình 2.3:** Minh họa cho dữ liệu raster có độ phân giải khác nhau

Độ phân giải là khả năng hiển thị đối tượng dưới dạng dữ liệu raster. Độ phân giải phụ thuộc vào kích thước của pixel. Kích thước của pixel càng nhỏ thì độ phân giải càng cao, đối tượng được hiển thị càng trung thực. Trong các hình trên, độ phân giải thay đổi theo kích thước của pixel cũng như số lượng pixel trong ảnh.

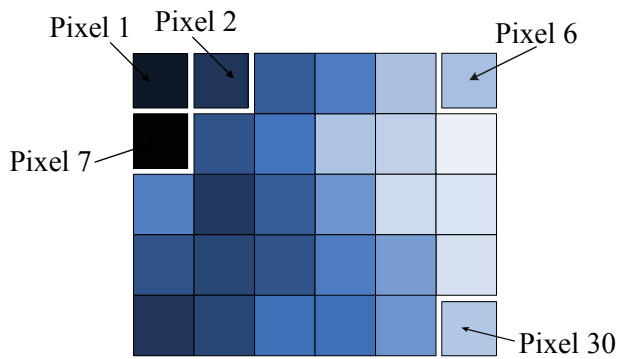
#### ➤ **Lưu trữ pixel**

Giống như mô hình dữ liệu vector, mô hình dữ liệu raster có thể biểu diễn được các đối tượng điểm, đường, vùng. Một điểm được biểu diễn như một giá trị trong ô độc nhất, một đường như một loạt các ô kết nối mô tả chiều dài, một vùng như một nhóm các ô kết nối mô tả hình dạng.

Độ chính xác của bản đồ phụ thuộc vào tỷ lệ bản đồ. Trong mô hình dữ liệu raster, độ phân giải và độ chính xác phụ thuộc vào cùng thế giới thực được biểu diễn bởi mỗi ô lưới. Vùng biểu diễn càng rộng, độ phân giải của dữ liệu càng thấp. Vùng được phủ càng nhỏ, độ phân giải càng lớn và các đối tượng càng được biểu diễn chính xác.

Do mô hình dữ liệu raster là một lưới đều nên các quan hệ không gian không được thể hiện rõ. Do vậy, không yêu cầu phải lưu trữ một cách rõ ràng các quan hệ không gian như đối với mô hình dữ liệu vector.

Đề ý rằng trong một lưới, mỗi ô có 8 ô láng giềng, 4 ở ngoài góc và 4 ở các bên (trừ các ô ở các mép ngoài). Các ô được nhận biết bởi vị trí của chúng ở trong lưới.



Số thứ tự Pixel	Giá trị Pixel
1	12
2	14
3	17
...	
15	17
...	
29	27
30	36

**\* Ưu điểm của cấu trúc dữ liệu raster:**

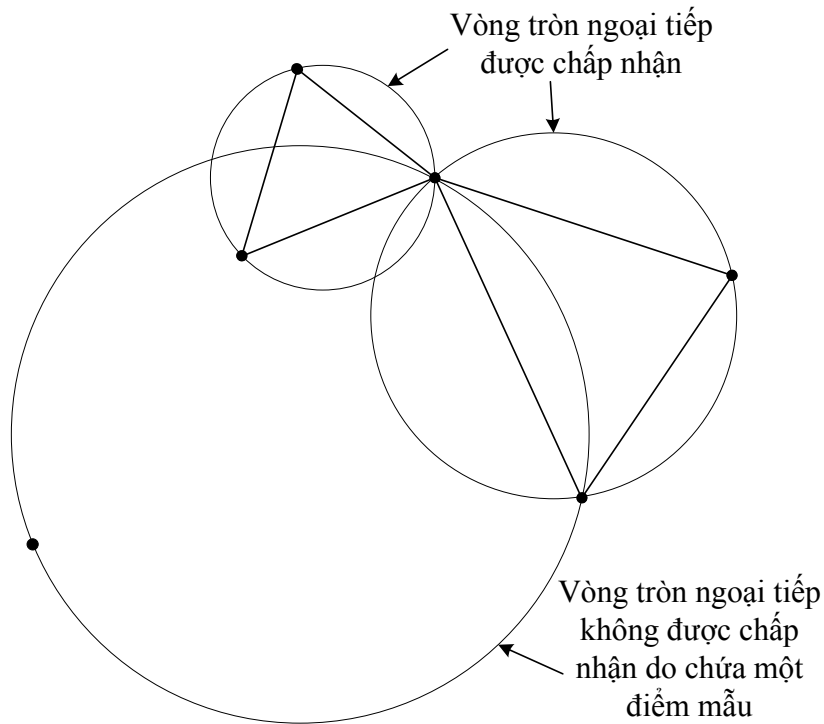
- Cấu trúc đơn giản, đồng nhất.
- Dễ chồng ghép bản đồ với các dữ liệu viễn thám.
- Dễ phân tích không gian, đặc biệt là không gian liên tục.
- Dễ mô hình hóa.

**\* Nhược điểm của cấu trúc dữ liệu raster:**

- Cần nhiều bộ nhớ.
- Khi giảm độ phân giải để giảm khối lượng dữ liệu sẽ làm giảm độ chính xác hay làm mất thông tin.
- Khó biểu diễn các mối quan hệ không gian.
- Không thích hợp với phân tích mạng.
- Đồ họa không đẹp.

### 2.3.3. Cấu trúc dữ liệu TIN

Lưới tam giác không đều (TIN - Triangulated Irregular Network) là một mô hình dữ liệu thường được sử dụng để mô tả độ cao địa hình. Các tọa độ x, y và z của các điểm được đo đạc và nhập vào mô hình dữ liệu TIN một cách đặc trưng. Các điểm này được phân bố trong không gian và có thể được kết nối với nhau theo kiểu hình thành các tam giác kích thước nhỏ nhất của bất kỳ ba điểm nào đó. Mô hình TIN được hình thành trên cơ sở mạng lưới các tam giác được kết nối với nhau. Các tam giác được tạo thành sao cho cạnh của tam giác này không cắt bất kỳ cạnh nào của các tam giác khác. Các đường cắt nhau được tránh bằng cách nhận dạng các đường tròn ngoại tiếp của bộ ba điểm. Đường tròn ngoại tiếp được định nghĩa là đường tròn đi qua cả ba điểm là đỉnh của một tam giác. Một tam giác chỉ được hình thành khi đường tròn ngoại tiếp của nó không chứa bất kỳ một điểm nào bên trong. Và mỗi tam giác sẽ xác định một mặt phẳng độ cao của cả mô hình.



Mô hình TIN thường sử dụng cách lập các quan hệ và kết nối các điểm lân cận với nhau. Mỗi cạnh của tam giác kết nối hai điểm và các điểm này lần lượt kết nối với các cạnh khác. Các kết nối này tiếp tục lan rộng cho đến khi mạng lưới được mở rộng tối đa. Như vậy, TIN là một mô hình phức tạp hơn so với các lưới raster đơn giản khi đối tượng là sự mô tả địa hình. Tuy nhiên, nó tỏ ra thích hợp hơn và hiệu quả hơn khi tiến hành mô tả dữ liệu địa hình của một vùng có yếu tố độ cao biến đổi.

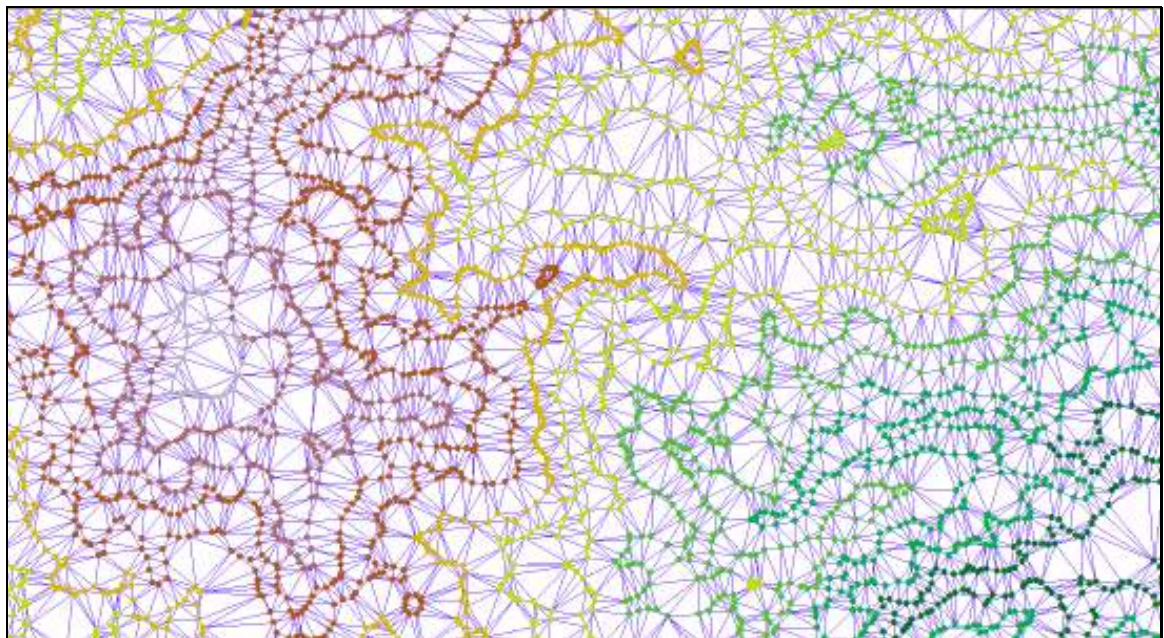
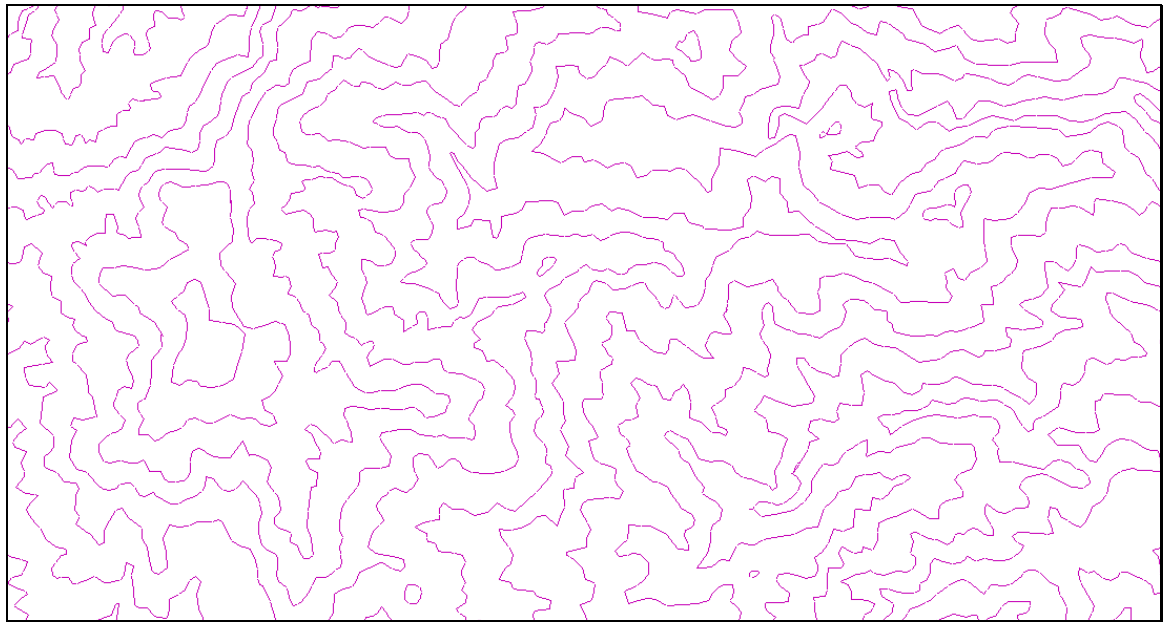
Xét một cách tương đối, chỉ cần dữ liệu của một số điểm là đủ cho việc mô tả các vùng rộng lớn hay các vùng liên tục, phẳng hay trơn tru. Nhưng khi mô tả các địa hình không liên tục và biến đổi mạnh thì đòi hỏi phải có dữ liệu về nhiều điểm hơn. Người lập bản đồ thường phải thu thập nhiều điểm mẫu hơn trên một đơn vị diện tích đối với những vùng có địa hình biến đổi mạnh.

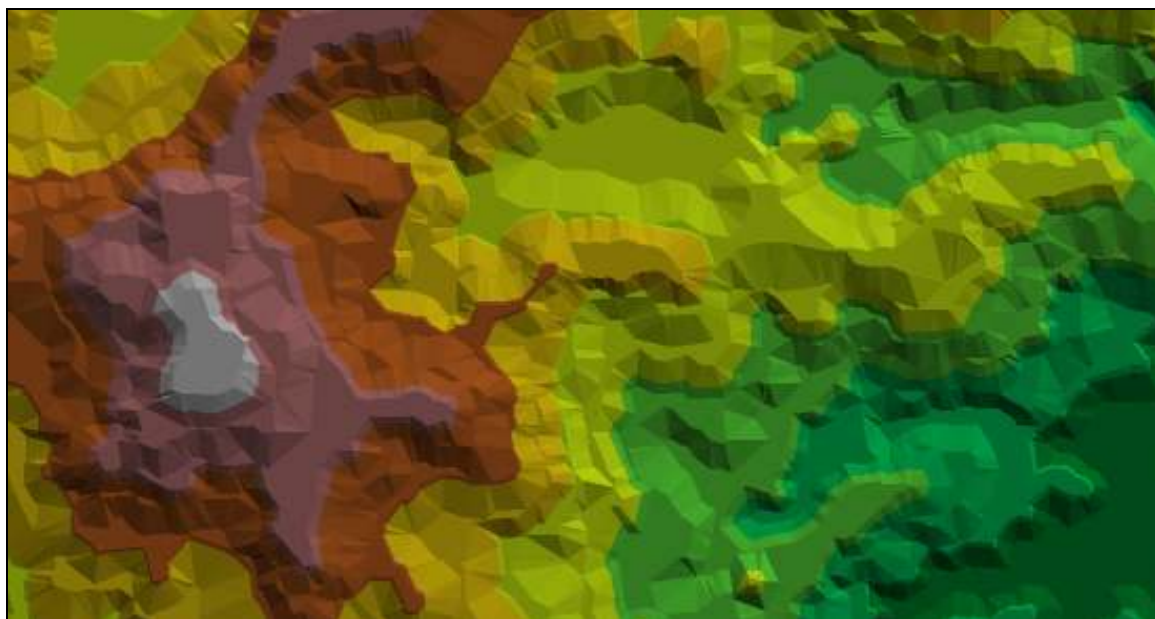
### **\* Ứng dụng TIN:**

Trong bản đồ học thì phương pháp truyền thống để biểu diễn bề mặt địa hình là đường bình độ. Tuy nhiên, đường bình độ không thuận tiện cho mục đích phân tích. Nếu có dữ liệu đường bình độ thì thông thường được chuyển sang một trong hai dạng biểu diễn địa hình chung nhất của GIS là lưới đều hay TIN.

Khoảng cách không đều của các điểm độ cao dẫn tới ta có tập các tam giác có kích thước và hình dáng khác nhau. Nơi các điểm dữ liệu gần nhau thì vùng nghiên cứu sẽ thay đổi nhanh độ cao. Nơi điểm dữ liệu xa nhau như trên mặt phẳng thì kích thước tam giác tăng nhanh. GIS chứa dữ liệu độ cao trong khuôn mẫu TIN cho phép tính toán độ dốc rất hiệu quả. Chúng cho phép phát sinh đường bình độ hay phác họa ảnh vùng nghiên cứu.







*Hình 2.4: So sánh đường bình độ với mô hình TIN điểm, đường và mô hình TIN bề mặt*

## **2.4. Cấu trúc cơ sở dữ liệu**

Cấu trúc cơ sở dữ liệu đề cập đến cách thức tổ chức các file dữ liệu trong một cơ sở dữ liệu. Khái niệm cơ sở dữ liệu là trọng tâm của GIS và là sự khác nhau chủ yếu giữa GIS và các hệ thống tạo bản đồ trên máy tính khác. Tất cả các hệ GIS hiện nay đều kết hợp chặt chẽ với hệ quản trị cơ sở dữ liệu.

Một cơ sở dữ liệu GIS hoàn chỉnh bao gồm:

- Cơ sở dữ liệu không gian
- Cơ sở dữ liệu phi không gian

Các cơ sở dữ liệu trên bao gồm các file máy tính chứa các dữ liệu về vị trí và dữ liệu mô tả về các đối tượng trên bản đồ. Mặt mạnh của một hệ GIS phụ thuộc vào khả năng liên kết hai kiểu dữ liệu này và duy trì được mối quan hệ không gian giữa các đối tượng bản đồ.

Khả năng tích hợp dữ liệu cho phép tìm kiếm và phân tích dữ liệu một cách hiệu quả. Người dùng có thể truy cập dữ liệu bảng thông qua bản đồ hoặc có thể tạo ra được bản đồ thông qua các cơ sở dữ liệu bảng. Để truy cập và hiển thị dữ liệu này, máy tính phải lưu trữ cả dữ liệu dạng bảng và dữ liệu đồ họa theo khuôn dạng có tổ chức và có thể tìm kiếm được.

### **2.4.1. Cơ sở dữ liệu không gian**

Cơ sở dữ liệu không gian bao gồm các file dữ liệu không gian. Như đã trình bày ở phần trước, trong GIS có hai mô hình dữ liệu không gian là vector và raster. Từ chính hai mô hình đó lại có các cấu trúc khác nhau. Điều này có nghĩa là sau khi nhập, ta được các dữ liệu thô (các cặp tọa độ hay các pixel). Các dữ liệu thô đó cần được cấu trúc lại để tạo thành các file dữ liệu trong cơ sở dữ liệu không gian trước khi sử dụng.

### ❖ **Mô hình dữ liệu vector**

Sau khi nhập dữ liệu không gian bằng các kỹ thuật khác nhau, ta được các file tọa độ với một tổ chức không gì khác ngoài tổ chức tuần tự tức là theo trình tự các file tạo ra. Các dữ liệu không gian đó phải được cấu trúc lại để biểu diễn và quản lý các đối tượng địa lý. Đối với dữ liệu vector, có ba loại cấu trúc được sử dụng phổ biến:

- Cấu trúc toàn vùng
- Cấu trúc spaghetti
- Cấu trúc topology

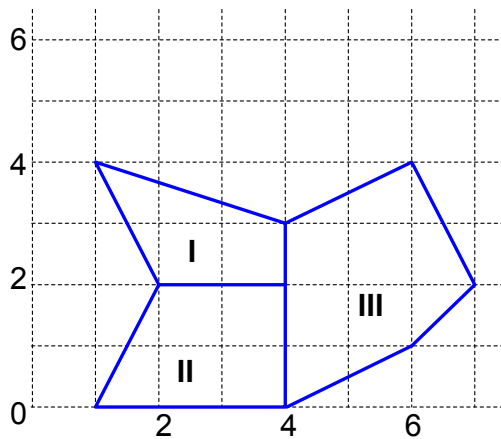
#### ➤ ***Cấu trúc toàn vùng***

Đây là dạng cấu trúc trong đó mỗi lớp của cơ sở dữ liệu được chia thành một nhóm vùng. Mỗi vùng được mã hóa thành trật tự các vị trí hình thành đường biên của vùng khép kín theo một hệ trục tọa độ nào đó.

- Mỗi đoạn thẳng xác định vùng đều được ghi lại hai lần.
- Một số điểm tạo nên cạnh của vùng được ghi lại nhiều lần

Ưu điểm của cấu trúc này là rõ ràng, dễ quản lý. Tuy nhiên, nó có nhược điểm là công kênh, khó cập nhật và chỉnh sửa dữ liệu.

Trong cấu trúc toàn đa giác thì các đoạn xác định đa giác được lưu lại hai lần trong cơ sở dữ liệu; các điểm tạo nên đa giác sẽ được lưu nhiều lần do đó việc cập nhật, sửa đổi dữ liệu trong tổ chức dữ liệu không gian loại này là rất khó khăn.



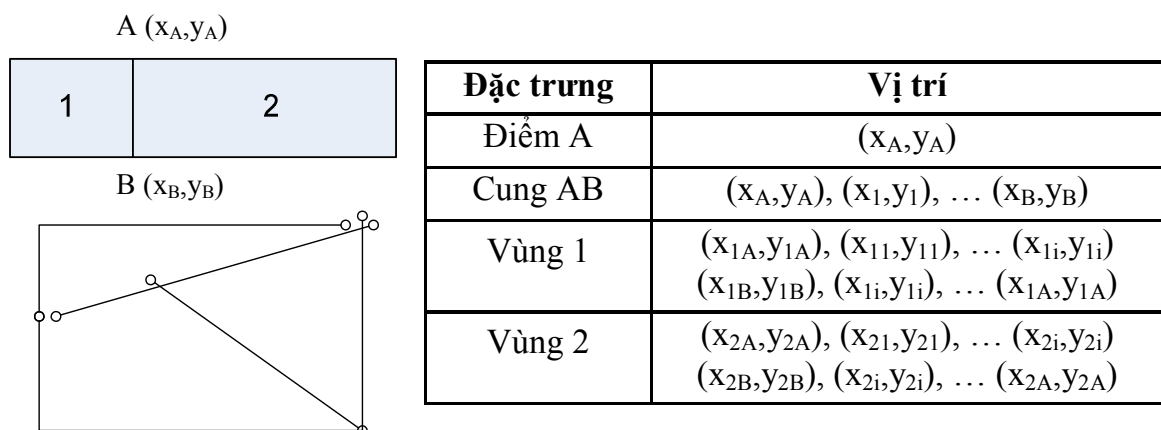
Vùng I	Vùng II	Vùng III
1,4	2,2	6,4
4,3	4,2	7,2
4,2	4,0	6,1
2,2	1,0	4,0
		4,3

➤ **Cấu trúc dữ liệu Spaghetti**

Đây là dạng cấu trúc sơ đẳng của dữ liệu vector trong đó mỗi đối tượng địa lý được mô tả bằng các thực thể hình học độc lập được biểu diễn bằng tọa độ hoặc bằng các phương trình tham số (đường thẳng, đường cong, đường tròn...).

Cấu trúc này rất hữu hiệu đối với công việc thiết kế và trình bày đồ họa song lại rất hạn chế đối với việc nghiên cứu các quan hệ giữa các đối tượng địa lý vì mỗi đối tượng độc lập với các đối tượng láng giềng.

Dữ liệu spaghetti thường được tạo ra từ việc số hóa thủ công các bản đồ trong đó ranh giới chung của các đa giác bị lặp lại do phải số hóa hai lần, dẫn đến dư thừa dữ liệu, tốn bộ nhớ và các cung có thể vắt qua nhưng không hề cắt nhau. Như vậy, dữ liệu spaghetti là một tập hợp các điểm và đường không có kết nối. Việc lưu trữ và tìm kiếm dữ liệu này là tuần tự và rất mất thời gian.



**Hình 2.3:** Cấu trúc Spaghetti

➤ **Cấu trúc dữ liệu Topology**

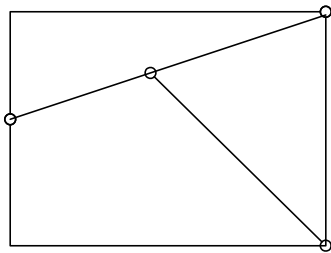
Trong GIS, topology được dùng để ghi lại và xử lý các mối quan hệ không gian giữa các đối tượng địa lý.

Một số thuật ngữ liên quan đến topology là nút, cung và vùng; trong đó nút là điểm đầu và điểm cuối của một cung và là điểm giao nhau của hai hay nhiều cung; cung là tập hợp các điểm kết nối với nhau và mỗi cung có một điểm đầu và điểm kết thúc; vùng là một đa giác khép kín được tạo thành bởi các cung.

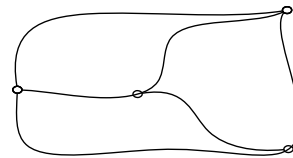
Các mối quan hệ không gian giữa các đối tượng địa lý được đề cập đến trong GIS là các mối quan hệ liền kề, tiếp nối và chứa đựng.

Trong các bản đồ số, các mối quan hệ không gian giữa các đối tượng được mô tả bằng cách sử dụng topology. Topology giúp xác lập rõ ràng các mối quan hệ không gian giữa các đối tượng độc lập với tọa độ của chúng.

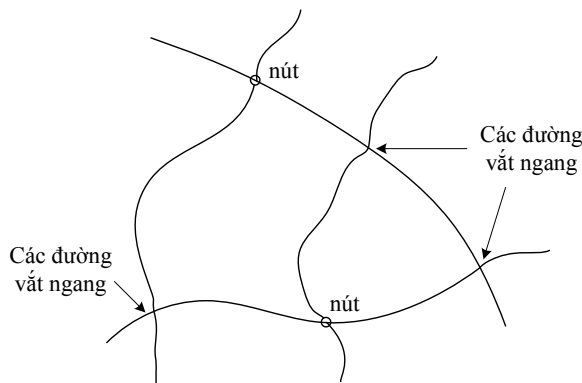
Việc tạo ra và lưu trữ các quan hệ không gian giữa các đối tượng địa lý có một số ưu điểm như dữ liệu sẽ được lưu trữ đầy đủ hơn khi sử dụng topology. Dữ liệu dư thừa được loại bỏ vì một cung có thể là một đối tượng tuyến hay một phần ranh giới của một đối tượng vùng hay cả hai. Vì vậy, ta có thể xử lý các dữ liệu nhanh chóng hơn và trên các tập dữ liệu lớn hơn. Khi tồn tại các quan hệ hình học, chúng ta còn có thể thực hiện các thao tác phân tích như tổ hợp các vùng kế cận có các đặc tính tương tự, chòong ghép các đối tượng địa lý.



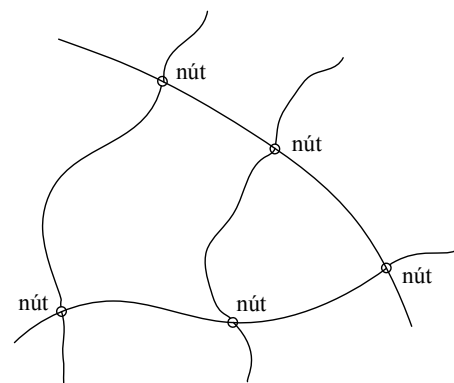
Cấu trúc topology



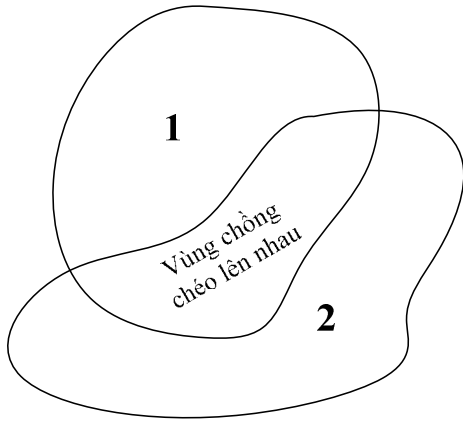
Cấu trúc topology biến dạng



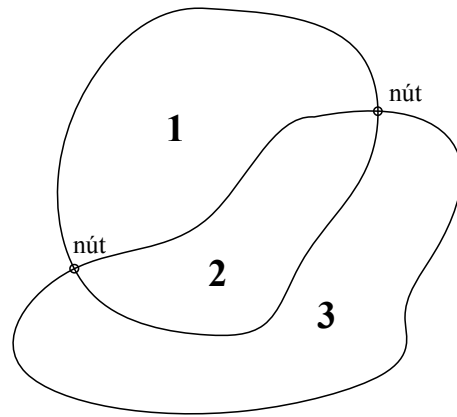
Cấu trúc topology đường không có mặt phẳng



Cấu trúc topology đường có mặt phẳng



Cấu trúc topology vùng không có mặt phẳng



Cấu trúc topology vùng có mặt phẳng

**Hình 2.5:** Cấu trúc dữ liệu topology

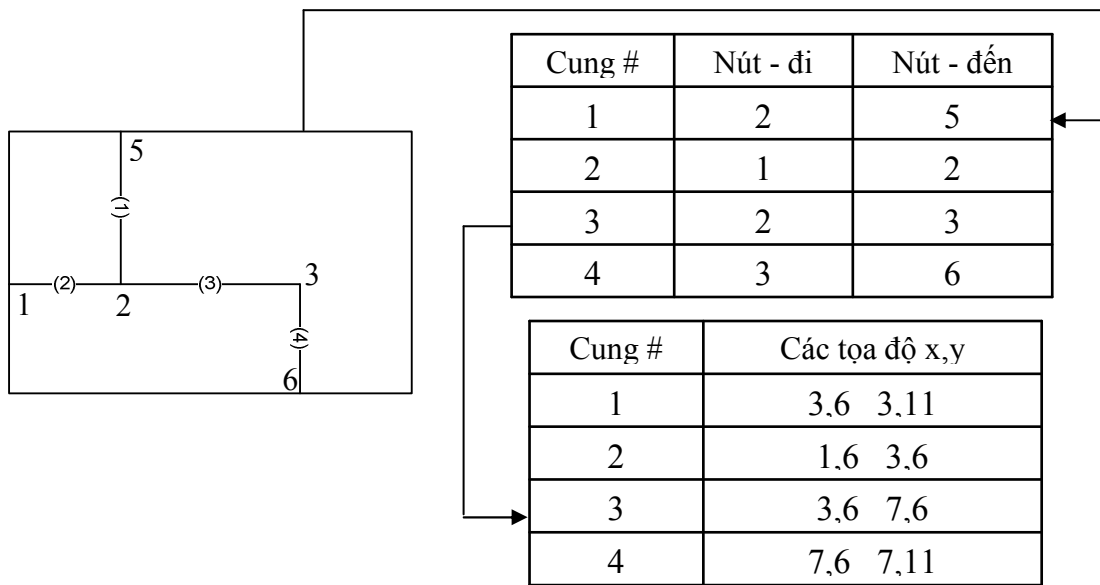
Có ba kiểu topology chủ yếu là topology cung - nút, vùng - cung và trái - phải. Cụ thể là:

- Các cung kết nối với nhau tại các nút (cung - nút) dùng để nhận biết mối liên kết giữa các đường.
- Các cung kết nối xung quanh để định nghĩa một vùng (vùng - cung) dùng để xác định một vùng.
- Các cung có hướng và kề cận trái, phải (trái - phải) dùng để nhận biết các vùng kề cận nhau.

✓ Topology cung - nút:

Các điểm có tọa độ  $(x,y)$  nằm dọc theo các cung sẽ xác định hình dạng của cung đó. Các điểm cuối của cung được gọi là các nút. Mỗi cung có hai nút: nút đi và nút đến. Các cung chỉ có thể nối với nhau tại các nút. Bằng cách đó, tất cả các cung gặp nhau tại một nút ta có thể biết được những cung nào được nối với nhau.

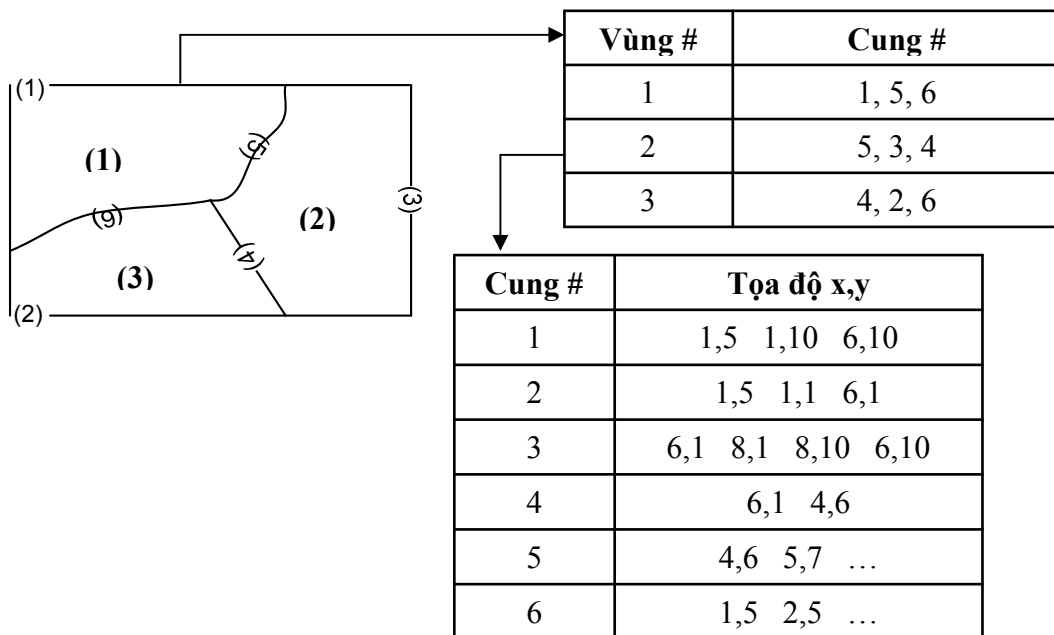
Ví dụ:



✓ Topology vùng - cung

Các vùng được biểu diễn bằng một dãy tọa độ (x,y) liên kết bao quanh một miền đồng nhất; một số hệ GIS lưu trữ theo khuôn dạng này. Tuy nhiên, lưu trữ tọa độ các cung xác định một vùng có thể sẽ có ích hơn là lưu trữ một tập các tọa độ (x,y). Danh sách các cung bao quanh một vùng cung được lưu trữ và sử dụng để cấu trúc thành các vùng khi cần thiết.

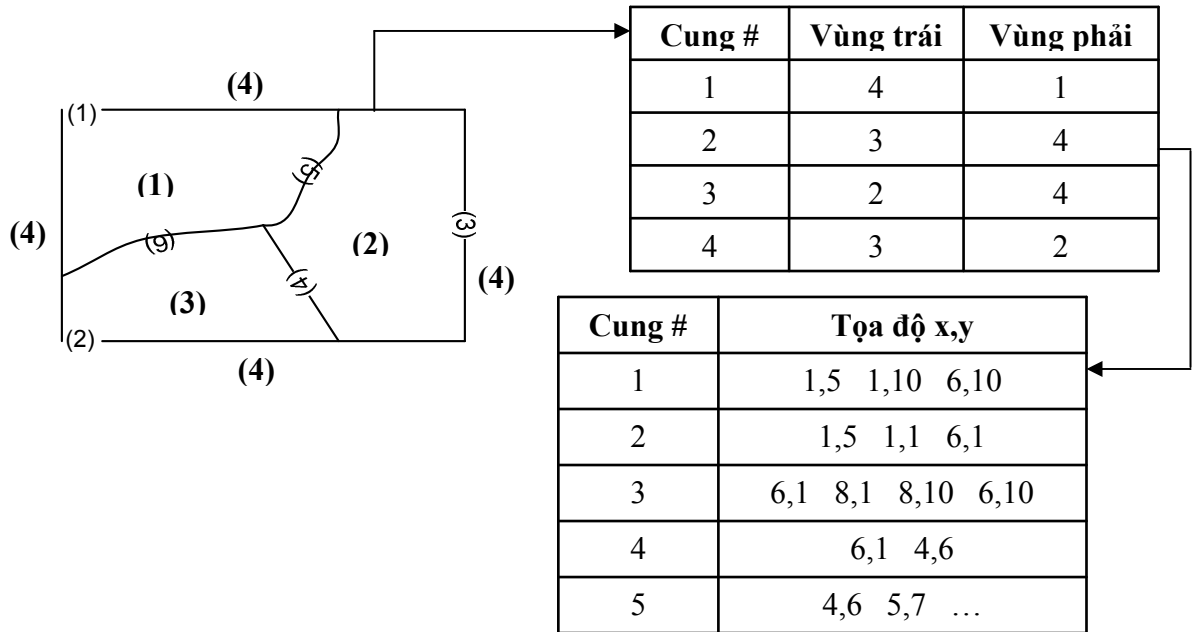
Ví dụ:



✓ Topology trái - phải

Bởi vì tất cả mọi cung đều có hướng (một nút đi và một nút đến) nên trong GIS sẽ lưu lại danh sách các vùng ở kê trái hoặc kê phải của một cung. Vì vậy, các vùng có chung một cung sẽ nằm kề nhau.

Ví dụ:



Bên cạnh các ưu điểm đã nêu trên, mô hình dữ liệu topology còn có một số nhược điểm như thời gian tính toán để nhận biết tất cả các nút lâu, dễ phát sinh lỗi liên quan đến việc khép kín các đa giác và tạo nút trong các mạng phức tạp và mỗi khi đưa dữ liệu mới vào, cập nhật dữ liệu. Các nút mới phải được tính toán và cập nhật các bảng topology mặc dù là trong các vector GIS thường có sẵn các chương trình để xây dựng và cập nhật topology.

❖ **Mô hình dữ liệu raster**

Với mô hình này, các file dữ liệu được tạo ra thường lớn do có sự lặp lại của các thông tin giống hệt nhau. Để khắc phục nhược điểm đó cần phải dùng các kỹ thuật nén dữ liệu qua đó giảm lượng dữ liệu và có nghĩa là giảm yêu cầu về không gian đĩa để lưu trữ dữ liệu.

Kỹ thuật nén file dữ liệu raster:

- Mã hóa theo dòng
- Mã hóa theo kiểu chia bốn
- ✓ Mã hóa theo dòng



Đây là kỹ thuật nén dữ liệu theo một chiều đối với file dữ liệu raster trong đó các ô liền nhau có cùng giá trị trên một dòng được nhóm lại. Kỹ thuật này có hiệu quả nhất khi gặp các diện đồng nhất lớn và ít loại.

Ví dụ: giả sử một ảnh có chứa một dòng với các giá trị ô như sau:

44444555556666662222233333

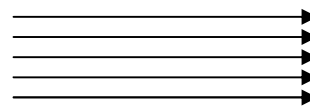
Kết quả mã hóa theo dòng đối với dòng này sẽ là: 54 55 66 52 53. Như vậy, có 5 bốn, 5 năm, 6 sáu, 5 hai và 5 ba.

Mã này rất hữu hiệu khi thuộc tính của lớp dữ liệu ảnh ít thay đổi. Ngược lại, nếu giá trị thuộc tính thay đổi nhiều, sự hữu hiệu của mã không cao mà đôi khi còn làm tăng dung lượng dữ liệu.

Mã hóa theo dòng tuy có thuật toán nén và giải mã đơn giản nhưng tỉ số nén không cao, thao tác tìm kiếm thường chậm vì lưu trữ theo hàng, không lưu ý đến tính lân cận.

	0	1	2	3	4	5	6	7	8	9
0	A	A	A	A	A	A	A	A	A	A
1	A	A	A	A	A	A	A	A	A	A
2	A	A	A	A	B	B	B	B	B	B
3	A	A	A	B	B	B	B	B	B	B
4	D	D	D	D	B	B	B	B	B	B
5	D	D	D	D	D	B	B	B	B	B
6	D	D	D	D	D	C	C	C	C	C
7	D	D	D	D	D	C	C	C	C	C
8	D	D	D	D	D	C	C	C	C	C
9	D	D	D	D	D	C	C	C	C	C

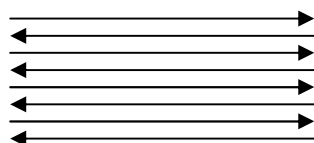
Hướng đọc các pixel theo mã hóa theo dòng:



Biểu diễn mã hóa theo dòng của lớp dữ liệu ảnh raster hình bên:

24A 6B 3A 7B 4D 6B 5D 5B 5D 5C 5D 5C  
5D 5C 5D 5C

Mã hóa theo dòng còn được cải biên theo cách đọc dữ liệu khác:



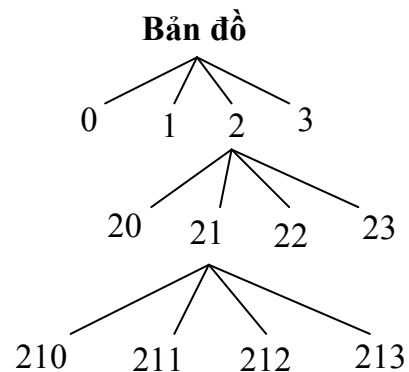
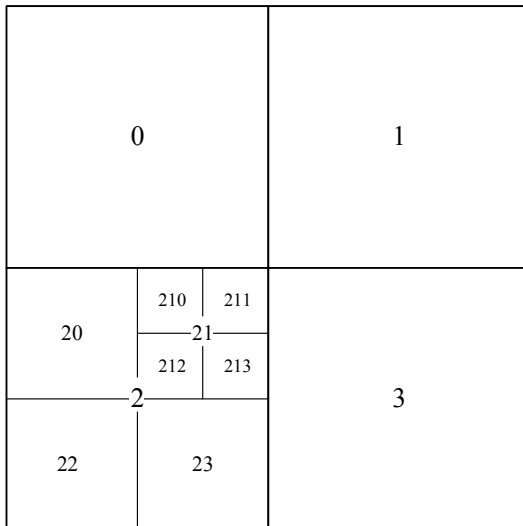
Kết quả mã hóa như sau:

24A 13B 3A 13B 4D 11B 10D 10C 10D 10C 5D

✓ Mã hóa theo kiểu chia bốn

Theo kỹ thuật này, không gian được chia nhỏ thành bốn phần bằng nhau từ khi bắt gặp vùng không đồng nhất. Việc chia nhỏ được tiếp tục cho đến các điểm nút cuối cùng tức là khi chỉ còn các ô vuông nguyên sinh có giá trị chuyên đề đồng nhất.

Với kỹ thuật này, có thể tiết kiệm được bộ nhớ bằng cách dùng các ô lớn hơn (có độ phân giải thấp hơn) để biểu diễn các vùng đồng nhất và các ô nhỏ hơn (có độ phân giải cao) cho những vùng cần chi tiết hóa.



Mức của cây			Thuộc tính
1	2	3	
0			Công nghiệp
1			...
2	20		Nhà ở
	21	210	Dịch vụ
		211	Thương mại
		212	Khu công cộng
		213	Khu giải trí
	22		Biệt thự
23		...	
3			Nông thôn

### 2.4.2. Cơ sở dữ liệu phi không gian

Cơ sở dữ liệu phi không gian bao gồm các file dữ liệu mô tả các đối tượng địa lý. Cũng như các dữ liệu không gian, các dữ liệu phi không gian hay còn gọi là các dữ liệu thuộc tính cũng phải được cấu trúc sao cho dễ quản lý và khai thác. Các thuộc tính mô tả được lưu trữ trong máy tính hoàn toàn tương tự như lưu trữ các giá trị tọa độ. Các thuộc tính được lưu trữ như một tập hợp các số và ký tự. Ví dụ như các thuộc tính của các đường phố bao gồm loại đường (cao tốc, quốc lộ, một chiều...), chất liệu (bê tông, nhựa đường, sỏi...), chiều rộng, số làn xe, tên... Các thông tin mô tả cho mỗi đoạn đường được lưu trữ trong máy tính thành một chuỗi giá trị trong khuôn dạng đã xác định trước, như thứ tự sau:

3 bê tông 40 4 Đào Tấn

Giá trị này sẽ tương đương cho một đường quốc lộ có mã hiệu là 3, được lát bằng bê tông, rộng 40m, có 4 làn và tên là Đào Tấn.

Trong GIS, cơ sở dữ liệu thuộc tính thường bao gồm một số lớn các file. Các file dữ liệu thường được tổ chức theo ba kiểu cấu trúc sau:

- Phân cấp
- Mạng
- Quan hệ

Trong đó, kiểu quan hệ được dùng phổ biến nhất và được coi là hiệu quả nhất. Tuy nhiên, một xu thế mới hơn dùng các hệ quản lý cơ sở dữ liệu hướng đối tượng đã bắt đầu trong các hệ thông tin nói chung và GIS nói riêng để khắc phục các nhược điểm của ba kiểu cấu trúc trên.

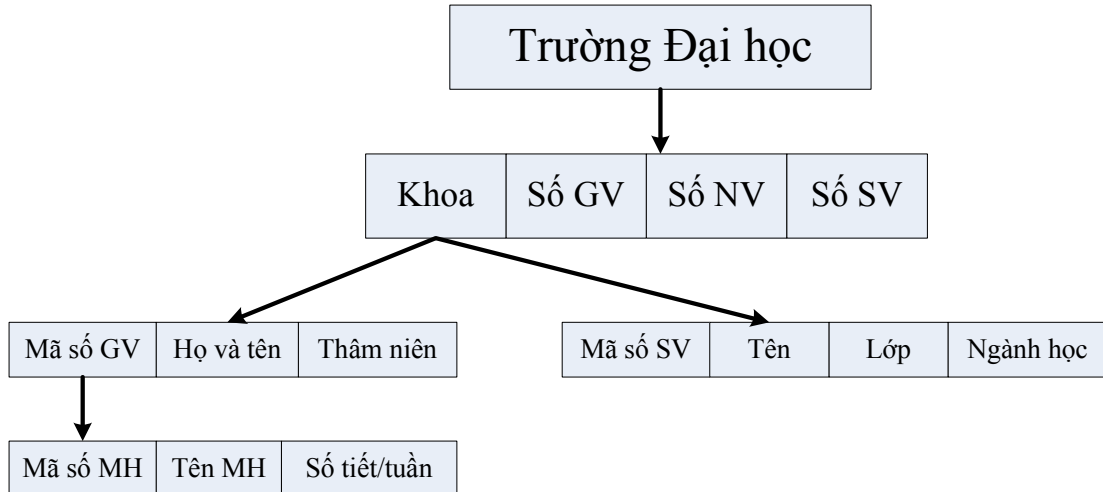
### ❖ Hệ quản lý cơ sở dữ liệu phân cấp

Trong hệ này, dữ liệu được phân loại theo một đẳng cấp được thiết lập rõ ràng. Cấu trúc này rất đơn giản, phù hợp cho một số kiểu quản lý, đặc biệt nếu biết trước các câu hỏi đặt ra. Song, nó bị hạn chế khi các dữ liệu không đồng nhất. Hiện mô hình này vẫn còn được sử dụng trên máy tính lớn và máy mini.

Trong mô hình phân cấp, dữ liệu được tổ chức theo cấu trúc cây. Tổ chức phân cấp xác định quan hệ giữa các thực thể và mã hóa chúng trong các bản ghi dữ liệu. Đỉnh của cây phân cấp gọi là gốc của cây được biểu diễn bởi một bản ghi dữ liệu có một hoặc nhiều trường.

Trừ phần tử gốc và các phần tử ngọn, mỗi phần tử trong cây đều có quan hệ với phần tử mức trên nó gọi là phần tử cha và một hoặc nhiều phần tử mức dưới nó gọi là phần tử con. Mỗi phần tử chỉ có một cha và có thể có nhiều con. Do đó, trong mô hình phân cấp, các quan hệ được thiết lập có thể là quan hệ một - một hay quan hệ nhiều - một.

Ví dụ:



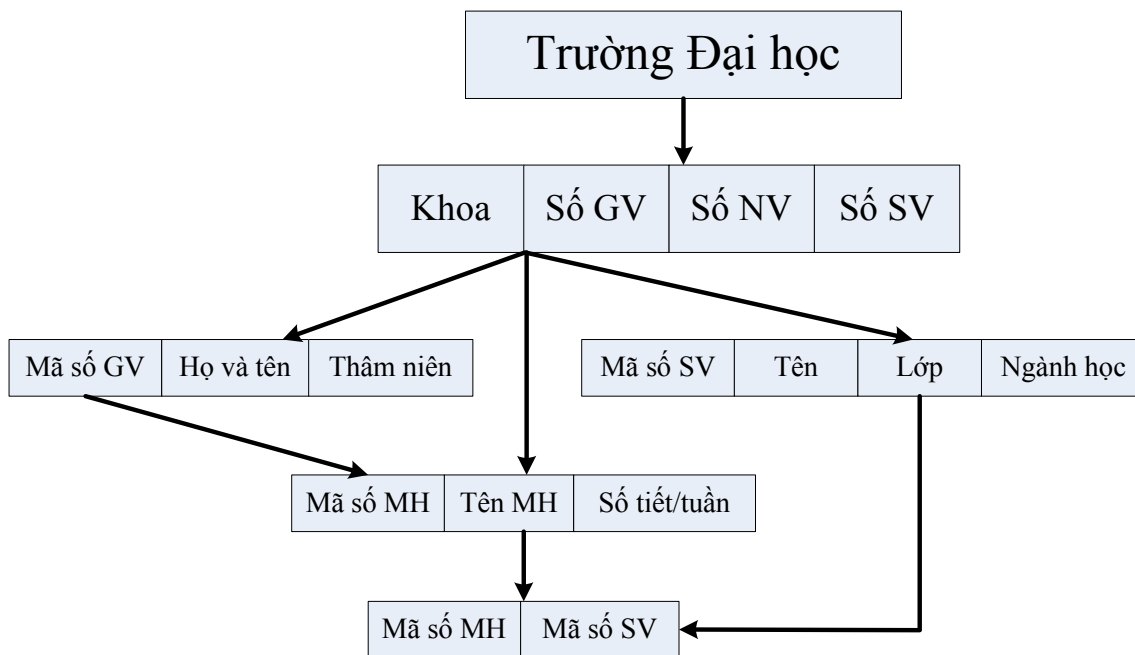
Trong mô hình phân cấp, việc truy cập được thực hiện qua cấu trúc cây, rất thuận lợi cho các bài toán tìm kiếm thông qua đường liên kết trực tiếp giữa hai thực thể. Trong ví dụ trên, ta có thể dễ dàng biết toàn bộ sinh viên và giáo viên trong một khoa vì có liên kết trực tiếp giữa thực thể sinh viên với thực thể khoa và thực thể giáo viên với thực thể khoa. Tuy nhiên, nếu muốn biết về số môn học trong khoa thì việc tìm kiếm sẽ khó khăn hơn vì không có liên kết trực tiếp giữa thực thể môn học và thực thể khoa mà phải thông qua thực thể giáo viên. Đây là một hạn chế của mô hình dữ liệu phân cấp. Hạn chế thứ hai của mô hình này là không thể tìm dữ liệu theo trường thuộc tính. Trong ví dụ trên, ta không thể tìm kiếm tất cả sinh viên thuộc một lớp nào đó hay học cùng một ngành nào đó do trường dữ liệu về lớp học hay ngành học không phải là trường khóa. Tuy nhiên, cấu trúc dữ liệu phân cấp có ưu điểm cơ bản là dễ hiểu và dễ cập nhật.

### ❖ Hệ quản lý cơ sở dữ liệu mạng

Mô hình mạng cung cấp các liên kết giữa các mục với nhau trong cơ sở dữ liệu. Cấu trúc này rất hiệu quả trong quản lý thông tin địa lý tuyến tính đặc biệt cho thiết lập topo mạng. Tuy nhiên, nó làm tăng tính phức tạp qua số con trỏ đưa vào giữa các cây. Cũng như mô hình phân cấp, hiện mô hình mạng vẫn còn được sử dụng trên các máy tính lớn và máy mini.

Mô hình mạng khắc phục sự thiếu mềm dẻo của mô hình phân cấp trong tiến trình truy vấn. Trong mô hình mạng, mỗi thực thể có thể có nhiều cha hoặc nhiều con và không nhất thiết phải có thực thể gốc. Do đó, khả năng tìm kiếm tốt hơn nhờ cách truy vấn trực tiếp các bản ghi, không qua các thực thể trung gian.

Ví dụ:



Cũng như mô hình phân cấp, mô hình mạng không chấp nhận quan hệ nhiều - nhiều. Do đó, một quan hệ trung gian như là một bảng giao có thể được thiết lập để mô tả quan hệ đó. Trong ví dụ trên, quan hệ sinh viên - môn học là một quan hệ nhiều - nhiều do mỗi sinh viên có thể đăng ký nhiều môn học và mỗi môn học có thể được đăng ký bởi nhiều sinh viên nên thiết lập bảng đăng ký là bảng giao để đảm bảo mỗi tổ hợp sinh viên - môn học là duy nhất. Mỗi thực thể môn học có nhiều bảng đăng ký và mỗi thực thể sinh viên cũng có nhiều bảng đăng ký. Như vậy, ta có hai quan hệ một - nhiều.

So với mô hình phân cấp, các thông tin về quan hệ trong mô hình mạng phức tạp hơn. Những thay đổi hoặc bổ sung quan hệ giữa các thực thể trong mô hình mạng khó thực hiện hơn so với trong mô hình phân cấp.

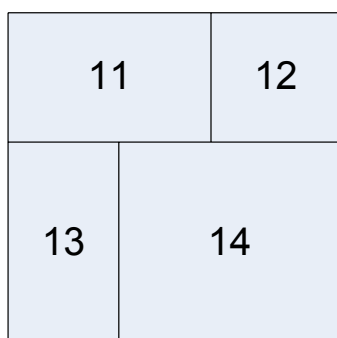
### ❖ Hệ quản lý cơ sở dữ liệu quan hệ

Đa số các hệ quản trị cơ sở dữ liệu hiện nay được thiết kế dựa trên mô hình quan hệ. Trong hệ này, dữ liệu được sắp xếp theo các bảng hai chiều chứa các bản ghi và các mối liên hệ của chúng. Ưu điểm của hệ này là rất linh hoạt và có thể trả lời mọi loại câu hỏi đặt ra bằng các toán tử logic hay các phép toán + - \* /.

Trong mô hình quan hệ, không có sự phân cấp các trường dữ liệu trong bản ghi. Mỗi trường dữ liệu có thể dùng như một khóa dữ liệu được lưu trữ như những bản ghi đơn giản gọi là bộ. Mỗi bộ biểu diễn một sự kiện, nhiều bộ được nhóm lại thành bảng hai chiều. Mỗi bảng được lưu trữ như một file riêng. Bảng dữ liệu biểu diễn toàn bộ các quan hệ giữa tất cả những thuộc tính trong nó. Bảng dữ liệu còn gọi là bảng quan hệ.

Trong mô hình quan hệ, sự tìm kiếm có thể thực hiện trên một bảng nào đó theo trường thuộc tính. Tìm kiếm các thuộc tính quan hệ lưu trữ trong nhiều bảng khác nhau được thực hiện bằng cách kết nối hai hay nhiều bảng quan hệ bằng toán tử “join”.

Ví dụ:



Mã vùng	Diện tích (ha)	Chu vi (m)	Mã vị
11	13	450	P-15
12	17	350	P-45
13	22	400	P-30
14	25	500	P-55

Mã vị trí	Chủ nhân	Tuổi
P-15	Nguyễn Văn Thái	50
P-45	Lê Thị Lan	45
P-30	Hà Thái Tâm	25
P-55	Phan Thị Thủy	30

**\* Ưu điểm của mô hình quan hệ so với mô hình phân cấp và mô hình mạng:**

- Mô hình quan hệ mềm dẻo hơn các mô hình khác, cách biểu diễn dữ liệu trong bảng quan hệ tạo điều kiện xử lý dễ dàng.
- Mô hình quan hệ có cơ sở toán học chặt chẽ cho phép áp dụng rộng rãi các công cụ đại số và logic.
- Tổ chức của mô hình quan hệ dễ hiểu.
- Một cách tổng quát, với cùng một cơ sở dữ liệu, biểu diễn theo mô hình quan hệ ít bị dư thừa như các mô hình khác.
- Mô hình quan hệ dễ đảm bảo tính an toàn dữ liệu vì có thể đặt mật khẩu truy nhập ở nhiều mức: mức quan hệ, mức thuộc tính, mức bộ, mức thuộc tính - bộ.

**\* Nhược điểm của mô hình quan hệ:**

- Mô hình quan hệ khó cài đặt hơn so với các mô hình khác.
- Hệ thống cơ sở dữ liệu quan hệ có khuynh hướng thao tác chậm hơn các hệ thống phân cấp hay hệ thống mạng.

### ***2.4.3. Kết nối các đối tượng và thuộc tính***

Thế mạnh của GIS là khả năng liên kết giữa các dữ liệu không gian và dữ liệu mô tả. Có ba đặc điểm đáng ghi nhớ của phép kết nối là:

- Mỗi quan hệ một - một giữa các đối tượng trên bản đồ và các bản ghi trong bảng thuộc tính đối tượng.
- Mỗi liên kết giữa đối tượng và bản ghi được duy trì thông qua khóa chung được gán cho mỗi đối tượng. Đối với các vùng khóa thì được gán bởi điểm nhãn của vùng đó.
- Khóa được lưu trữ vật lý vào hai nơi: trong các file chứa cặp tọa độ x,y và các bản ghi tương ứng trong bảng dữ liệu thuộc tính đối tượng.

Chú ý rằng cả hai bản ghi tọa độ và thuộc tính đều có phần tử chung là số hiệu đối tượng. Số hiệu này dùng để kết nối các thuộc tính với các tọa độ của đối tượng, duy trì liên kết một - một giữa các bản ghi tọa độ và bản ghi thuộc tính. Khi phép kết nối được thiết lập, ta có thể truy vấn bản đồ để hiển thị thông tin thuộc tính, hoặc tạo ra bản đồ dựa trên thông tin thuộc tính được lưu trữ trong bảng thuộc tính đối tượng.

## CHƯƠNG 3

---

# Nhập dữ liệu

### 3.1. Khái quát

Trong GIS, nhập dữ liệu là quá trình đưa các dữ liệu, có thể là dữ liệu không gian hay dữ liệu thuộc tính vào trong máy tính để tạo ra một cơ sở dữ liệu số mới hay bổ sung, cập nhật một cơ sở dữ liệu số đã có sẵn.

Có thể nhập dữ liệu không gian vào trong máy tính theo một trong các cách sau đây:

- Số hóa bản đồ bằng bàn số
- Quét bản đồ bằng máy quét
- Dùng bàn phím để đưa tọa độ vào
- Nạp dữ liệu số từ trên băng, trên đĩa vào máy tính
- Nhập dữ liệu từ các máy tính khác thông qua mạng

Tương tự, ta cũng có thể nhập dữ liệu thuộc tính vào trong máy tính bằng các cách khác nhau:

- Đưa dữ liệu thuộc tính vào trong máy tính từ bàn phím
- Nạp các file dữ liệu thuộc tính sẵn có trên băng hay đĩa
- Nhập các file sẵn có trên các máy tính khác thông qua mạng

### 3.2. Các công nghệ thu thập dữ liệu

#### 3.2.1. Số hóa

Số hóa là quá trình chuyển các dữ liệu tương tự sang dạng số bằng thủ công, tự động hay bán tự động. Song trong thực tế, người ta hay gọi số hóa thay cho số hóa thủ công và phân biệt nó với việc quét dữ liệu.

Số hóa là một công nghệ được sử dụng rộng rãi để sản xuất dữ liệu số từ dữ liệu tương tự. Các đối tượng điểm, đường, vùng hình thành nên bản đồ được chuyển thành các tọa độ  $x,y$ . Một điểm được thể hiện bằng một cặp tọa độ, một đường được thể hiện bằng một chuỗi các cặp tọa độ và khi được kết nối, một hoặc



nhiều đường với điểm nhãn bên trong đường biên sẽ xác định được một vùng. Do vậy, số hóa được xem như là một quá trình nhập các điểm, đường và vùng.

Các điểm được sử dụng với hai mục đích: thể hiện các đối tượng điểm hoặc xác định một vùng; vì vậy để tránh nhầm lẫn, người ta không số hóa cả hai loại điểm này trên cùng một lớp bản đồ.

Thiết bị số hóa phổ biến nhất được sử dụng để số hóa bản đồ bao gồm một bàn và một con trỏ. Bàn số hóa được phân biệt bởi hãng sản xuất, kích thước và độ phân giải. Mỗi bàn số bao gồm các sợi dây điện nhỏ chạy ngang và dọc bên trong tạo thành một lưới. Con trỏ bàn số có một bộ phận nhìn quang học có các sợi chữ thập cho phép người sử dụng định kiểm bằng mắt trên bản đồ. Con trỏ có thể là loại không dây hoặc có dây.



**Hình 3.1:** Bàn số hóa

Các nút trên con trỏ được lập trình để tiến hành một số chức năng, chẳng hạn như ghi lại một điểm hoặc bắt đầu và kết thúc một đường. Khi con trỏ được ấn, máy tính sẽ ghi lại các tọa độ x,y của vị trí hiện thời. Đây chính là các tọa độ của các đối tượng điểm hay một trong những điểm hợp thành đường hay vùng.

Để số hóa, bản đồ phải được gắn trên bàn số hóa, các điểm và các đường được vạch bằng con trỏ của bàn số. Vùng cảm ứng điện từ của bàn số không mở rộng đến các mép của nó vì vậy, để xác định các giá trị tọa độ, ta phải đảm bảo bản đồ được đặt trong vùng hoạt động.

Hai hay nhiều đường giao nhau tại một điểm thì được gọi là giao điểm hay điểm nút. Khi số hóa, ta có thể chọn các giao điểm như là các nút hoặc có thể bỏ đi các giao điểm. Cả hai phương pháp đều có thể chấp nhận được. Số hóa spaghetti được đặt tên như vậy bởi các giao điểm là không được xác định khi các cung được số hóa. Phương pháp này thường được sử dụng để xác định các đường spaghetti như các đường biên. Đối với các đối tượng chính xác hoặc các đối tượng cong và xoắn lại với nhau, ta nên số hóa riêng lẻ. Xác định các giao điểm sẽ nâng cao độ chính xác của tọa độ. Các giao điểm chưa được xác định trong quá trình số hóa thì sẽ được xác định sau đó trong quá trình tự động hóa.

Số hóa thủ công có thể được thực hiện theo hai thức: số hóa điểm và số hóa dòng. Trong số hóa điểm, người thao tác đặt con trỏ bàn số đến điểm cần số hóa và ấn nút để đưa tọa độ của điểm đó vào trong máy tính. Trong thức số hóa dòng, người thao tác không cần phải ấn nút để số hóa từng điểm một mà di chuyển con trỏ bàn số theo đối tượng cần số hóa và các điểm cần số hóa sẽ được đưa vào hành loạt.

Số hóa điểm có ưu điểm là chính xác, tiết kiệm được bộ nhớ và phù hợp với các đối tượng điểm, đoạn thẳng; tuy nhiên nó có nhược điểm là chậm. Số hóa dòng có ưu điểm là nhanh, phù hợp với các đối tượng đường cong, đường ngoằn ngoèo, các đường đồng mức; nhưng nhược điểm của nó là kém chính xác và tốn bộ nhớ để lưu trữ.

Thông thường, để đảm bảo số hóa một cách có hiệu quả và chính xác, quy trình số hóa phải đảm bảo được thực hiện theo các bước sau:

- Chọn dùng bản đồ gốc tốt.
- Xác định các thủ tục cần thiết như quy ước đặt tên chuẩn, các kế hoạch, các thay đổi hay các thủ tục chuẩn khác.
- Chuẩn bị bản đồ.
- Tiến hành số hóa bản đồ.
- Tìm kiếm và hiệu chỉnh lỗi.

### **3.2.2. Quét bản đồ**

Quét là một phương pháp số hóa tự động dường như không cần đến sự can thiệp của các thao tác viên, qua đó dữ liệu tương tự được chuyển sang dữ liệu số bằng thiết bị quét. Đó là phương pháp nhanh nhất để thu thập dữ liệu về hàng loạt các đối tượng như khi số hóa các ảnh hay các bản đồ sạch (bản đồ không có văn bản và các đường thừa).

Công nghệ quét dùng các thiết bị laser hay tương tự để chuyển dữ liệu tương tự sang dạng số. Các máy quét khác nhau ở hãng sản xuất, kích cỡ, kiểu dáng, độ phân giải, màu hay đen trắng.



**Hình 3.2:** Máy quét bản đồ

Kết quả quét là các file dữ liệu raster trong đó mỗi pixel mang một giá trị khác nhau. Chất lượng quét phụ thuộc vào độ chính xác của máy quét, tỷ lệ với số điểm trên một inch. Dữ liệu raster có thể được phân tích và chuyển sang dữ liệu vector bằng hai cách: vector hóa tự động và vector hóa thủ công (hay còn gọi là số hóa trên màn hình).

Quá trình quét bao gồm các bước như chuẩn bị, quét và biên tập dữ liệu. Công tác chuẩn bị kéo theo việc chuẩn bị phần cứng như máy tính, máy quét, phần mềm và dữ liệu cần quét.

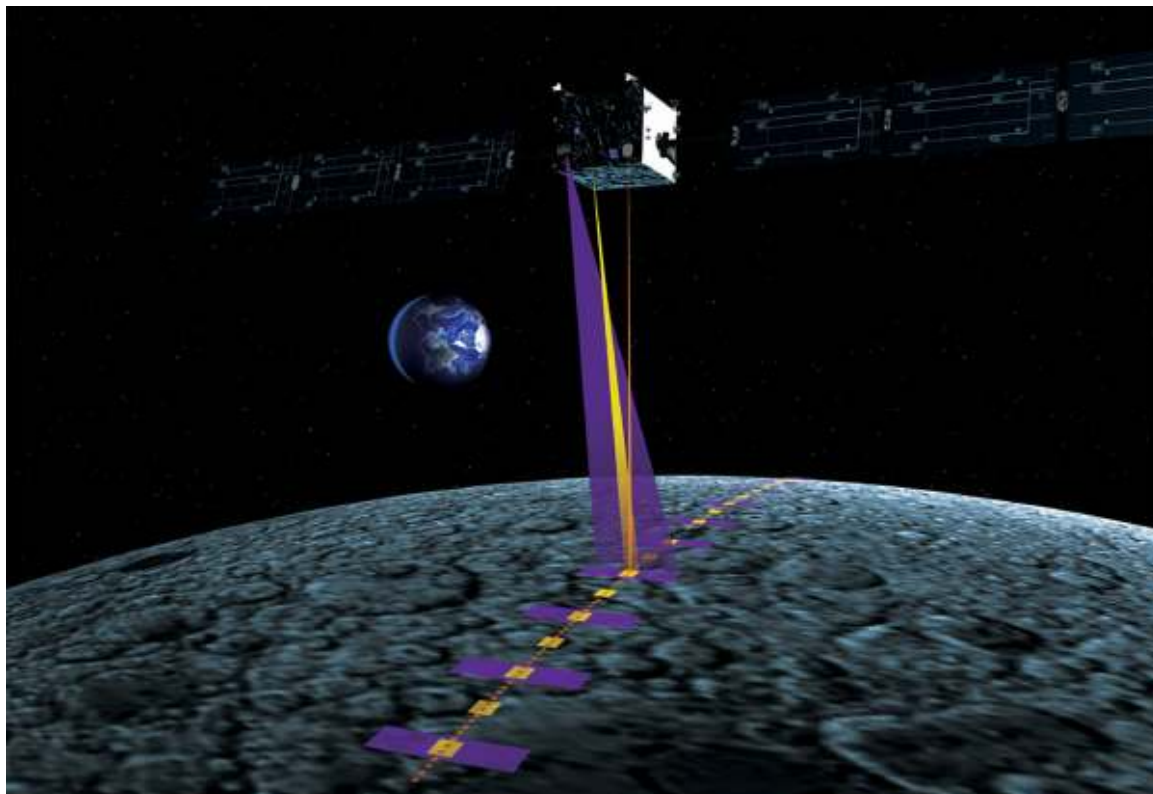
Như vậy, ta có hai cách để chuyển dữ liệu không gian ở dạng tương tự sang dạng số trong GIS là: số hóa thủ công và quét. Các ưu và nhược điểm của hai phương pháp này có thể được so sánh như sau: số hóa thủ công có ưu điểm là kỹ thuật tương đối đơn giản, tận dụng lao động và tiết kiệm bộ nhớ nhưng phương pháp này chậm, chủ quan và dễ bị lỗi. Phương pháp quét có ưu điểm là nhanh, khách quan nhưng tốn bộ nhớ và đòi hỏi phải lựa chọn kỹ dữ liệu tương tự. Sự khác biệt này là cơ sở cho việc lựa chọn sử dụng phương pháp nào trong từng trường hợp cụ thể.

### **3.2.3. Viễn thám**

Viễn thám là một công nghệ hiện đại được sử dụng để thu thập dữ liệu từ xa về các đối tượng, hiện tượng hay quá trình xảy ra trên bề mặt Trái đất mà không hề có sự tiếp xúc trực tiếp với các đối tượng, hiện tượng hay quá trình đó.

Cơ sở vật lý của phương pháp viễn thám là năng lượng điện từ truyền từ một nguồn nào đó tới các vật thể nằm trên bề mặt Trái đất được phản xạ, khúc xạ, phát xạ và hấp thụ bởi các vật đó. Căn cứ vào nguồn bức xạ điện từ sử dụng trong viễn thám, người ta phân biệt viễn thám bị động và viễn thám chủ động. Viễn thám bị động sử dụng ánh sáng mặt trời làm nguồn bức xạ điện từ do đó nó phụ thuộc vào điều kiện thời tiết và thời điểm trong ngày. Viễn thám chủ động hoạt động trong mọi điều kiện thời tiết kể cả ngày lẫn đêm do chủ động được nguồn năng lượng điện từ bằng việc sử dụng các bộ cảm biến chủ động như radar và lazer.

Ngoài ra, viễn thám còn được chia thành nhiều loại khác trên các cơ sở khác nhau như viễn thám hàng không và viễn thám vệ tinh, viễn thám chụp ảnh và viễn thám không chụp ảnh, viễn thám nhìn thấy, viễn thám cận hồng ngoại, viễn thám hồng ngoại nhiệt, viễn thám vi sóng và viễn thám dùng sóng radio.



*Hình 3.3: Thiết bị viễn thám SMART-1 đang quét bề mặt Mặt trăng*

Dữ liệu viễn thám rất đa dạng, là sản phẩm của nhiều loại viễn thám khác nhau và là nguồn dữ liệu lớn, quý báu để xây dựng cơ sở dữ liệu GIS. Tuy nhiên, căn cứ vào khuôn dạng dữ liệu, ta có thể phân chia tất cả các dữ liệu viễn thám thành hai loại:

- Dữ liệu tương tự: là dữ liệu dưới dạng bản cứng, gồm các bản in trên giấy, phim, các âm bản đen trắng hay màu ở tỷ lệ khác nhau. Dữ liệu tương tự có thể đọc và giải đoán bằng mắt thường.
- Dữ liệu số: là dữ liệu dưới dạng bản mềm được lưu trữ trong các môi trường tương thích như băng, đĩa từ, đĩa quang. Dữ liệu dạng này không thể đọc và giải đoán bằng mắt thường mà cần đến phần cứng, phần mềm chuyên dụng hợp thành một hệ thống xử lý ảnh số.

Không giống như các dữ liệu địa lý dạng vector, các dữ liệu viễn thám ở dạng raster trong đó từng diện tích nhỏ của bề mặt Trái đất được thể hiện thông qua một số đặc tính. Những dữ liệu đó được xử lý, phân tích bằng cách sử dụng hệ thống xử lý ảnh số được cấu thành từ các phần cứng, phần mềm đặc biệt. Các dữ liệu viễn thám được đưa vào GIS bằng cách số hóa thủ công hay quét các ảnh

tương tự và bằng cách nhập các file ảnh số. Các loại dữ liệu viễn thám vệ tinh được dùng để xây dựng cơ sở dữ liệu GIS có sự khác nhau ở nguồn cung cấp, độ phân giải, độ lặp lại và diện phủ.

Hiện nay, trên thế giới cũng như ở Việt Nam, có nhiều phần mềm xử lý ảnh số do các hãng khác nhau sản xuất và tiếp thị như DIDACTIM, ERDAS, EASI/PACE, IDRISI, ILWIS... Mỗi phần mềm đều có những điểm mạnh và điểm yếu riêng của nó do vậy việc lựa chọn và sử dụng tổng hợp các phần mềm sẵn có là một vấn đề cần được xem xét kỹ trong công tác xử lý ảnh số.

### **3.2.4. Đo đạc**

Công tác đo đạc được tiến hành trên mặt đất nhằm xác định trực tiếp tọa độ của các điểm trên bề mặt Trái đất bằng cách đo khoảng cách và góc, hướng của chúng so với các điểm đã biết (các điểm mốc) bằng các dụng cụ như thước, địa bàn và máy kinh vĩ. Hệ quy chiếu có thể là hệ quy chiếu địa phương, khu vực, quốc gia hay quốc tế.

Thiết bị đo đạc bao gồm các thiết bị quang cơ, điện tử và quang điện tử. Thiết bị đo đạc tự động như máy kinh vĩ toàn trạm cho phép xử lý tự động các dữ liệu đo đạc. Dữ liệu đo đạc có thể đưa vào GIS thông qua việc nhập từ bàn phím hay trực tiếp từ file máy tính.



*Hình 3.4: Đo đạc thực địa*

### **3.2.5. Hệ thống định vị toàn cầu**

Hệ thống định vị toàn cầu (GPS - Global Positioning System) là một công nghệ dựa trên nền các vệ tinh cung cấp thông tin về vị trí chính xác trong mọi điều kiện thời tiết, mọi thời điểm trong ngày và mọi điều kiện địa hình. Công nghệ GPS trợ giúp xác định vị trí và đường đi của các thuyền, máy bay, các phương tiện đường bộ lớn và nhỏ... và các thiết bị GPS nhỏ và nhẹ ngày được phát triển để có thể dễ dàng mang theo và sử dụng.

Cho đến năm 2004, đã có hai hệ thống GPS vệ tinh và một hệ thống thứ ba đang được nghiên cứu để xây dựng. Bộ quốc phòng Mỹ đang sử dụng hệ thống



NAVSTAR cho cả mục đích quân sự và dân sự. Một hệ thống của Nga có tên gọi là GLONASS cũng đang tồn tại nhưng ít được sử dụng, chủ yếu là sử dụng vào thời Liên Xô cũ. Hệ thống thứ ba có tên là GALILEO đang được nghiên cứu cho mục đích phi quân sự và được thiết kế và xây dựng bởi các chính phủ thuộc liên minh Châu Âu và các tập đoàn công nghiệp.

GPS có ba thành phần chính. Thứ nhất là vệ tinh. Thành phần này bao gồm một chùm vệ tinh có quỹ đạo bay xung quanh Trái đất ở độ cao so với mặt nước biển khoảng 20.000 Km. Hệ thống được thiết kế để hoạt động với 21 vệ tinh GPS và 3 vệ tinh dự phòng. Các vệ tinh này được phân bố trên 6 mặt phẳng quỹ đạo khác nhau. Tất cả các vệ tinh đều quay quanh Trái đất 2 vòng mỗi ngày và mỗi vệ tinh tồn tại trên đường chân trời trong vòng 8 tiếng hay nhiều hơn trong mỗi ngày. Bộ các vệ tinh hoạt động và dự phòng đã được lập kế trình trước và cả hai loại đều tồn tại lâu hơn so với thời gian tồn tại được thiết kế do đó thực tế tồn tại hơn 24 vệ tinh trên quỹ đạo trong cùng một thời điểm. Khoảng từ 4 đến 8 vệ tinh hoạt động có thể nhìn thấy được từ bất kỳ một vị trí nhìn trên Trái đất nào mà không bị chướng ngại.



**Hình 3.5:** Phân bố các vệ tinh trên mặt phẳng quỹ đạo xung quanh Trái đất

Thành phần thứ hai của GPS là bộ phận điều khiển. Bộ phận điều khiển có nhiệm vụ theo dõi đường bay, giao tiếp, thu dữ liệu, phân tích và điều khiển. Bộ phận này được sử dụng để quan sát, duy trì và quản lý các vệ tinh GPS cũng như cả hệ thống. Có 5 trạm theo dõi đường bay phân bố trên Trái đất trong đó trạm điều khiển chủ được đặt tại bang Colorado, Mỹ. Dữ liệu được thu thập thông qua một số nguồn bởi trạm điều khiển chủ. Các dữ liệu này bao gồm cả thông tin về tình trạng sức khỏe của mỗi vệ tinh và thông tin về đường bay của các vệ tinh. Trạm điều khiển chủ tiến hành tổng hợp các thông tin đó và đưa ra các phản hồi liên quan đến sự di chuyển, sự tính toán về mặt thời gian và các dữ liệu khác đến mỗi vệ tinh. Trạm điều khiển chủ cũng phát tín hiệu đến các vệ tinh về các chỉnh sửa trong tiến trình, các thay đổi trong hoạt động hay các thay đổi khác.



**Hình 3.6:** Một thiết bị thu GPS của hãng Garmin

Thành phần thứ ba của GPS đó là bộ phận người sử dụng. Bộ phận này bao gồm các cá nhân riêng lẻ hay nhóm người cùng với một hay nhiều thiết bị thu GPS. Thiết bị thu GPS là một thiết bị có khả năng ghi lại dữ liệu được truyền về từ vệ tinh. Có nhiều hãng khác nhau sản xuất thiết bị này như Garmin, Trimble, Leica... Thiết bị thu GPS có thể đeo ở người, cầm tay, gắn trên các phương tiện

giao thông. Dữ liệu định vị GPS thu được là các dữ liệu số do đó có thể nhập trực tiếp vào GIS.

### **3.2.6. Chuyển đổi dữ liệu**

Sự đa dạng về phần mềm máy tính nói chung và phần mềm GIS nói riêng đã dẫn đến sự đa dạng về khuôn dạng dữ liệu được dùng để xây dựng cơ sở dữ liệu địa lý. Điều đó cũng làm nảy sinh nhu cầu phát triển các chương trình chuyển đổi dữ liệu hay các chương trình nhập và xuất dữ liệu trong các hệ GIS.

Một hệ GIS mạnh cho phép nhập dữ liệu từ nhiều nguồn và ở các khuôn dạng khác nhau. Mỗi hệ GIS có một tổ chức dữ liệu đặc trưng và việc truyền dữ liệu từ hệ GIS này sang hệ GIS khác cần đến một giao diện thích hợp. Nếu dữ liệu ta đang cần đã có khuôn dạng số ở đâu đó thì có thể không cần số hóa từ bản đồ phác thảo mà có thể lấy từ các cơ quan của chính phủ hay các công ty thương mại. Việc chuyển đổi khuôn dạng dữ liệu được áp dụng đối với cả dữ liệu không gian và phi không gian.

Thời gian để chuyển đổi khuôn dạng đối với những dữ liệu sẵn có để tạo ra các lớp bản đồ có thể kéo dài từ vài phút đến vài giờ, thậm chí lâu hơn. Quá trình xử lý và thời lượng cần thiết phụ thuộc vào số lượng và chất lượng của dữ liệu được chuyển đổi và khả năng tương thích với các khuôn dạng khác như thế nào.

### **3.3. So sánh và lựa chọn phương pháp nhập dữ liệu**

Nhập dữ liệu là một bước quan trọng trong việc tạo mới, bổ sung và cập nhật cơ sở dữ liệu GIS. Do đó, trước khi tiến hành công tác nhập dữ liệu, cần tiến hành so sánh các phương pháp hiện có để tìm ra phương pháp thích hợp nhất với mục đích nhập và sử dụng dữ liệu GIS sau này. Có thể so sánh, đánh giá các phương pháp nhập dữ liệu khác nhau trên cơ sở phân tích các yếu tố sau:

- Thời lượng cần để chuẩn bị dữ liệu.
- Thời lượng cần để vào dữ liệu.
- Độ chính xác của dữ liệu và ảnh hưởng của nó đến các quá trình biên tập và xử lý dữ liệu như thế nào.

Khi lựa chọn phương pháp nhập dữ liệu, cần xem xét các điểm sau:

- Phần cứng và phần mềm có sẵn.
- Kinh nghiệm của các thao tác viên.
- Loại tài liệu gốc.
- Chất lượng dữ liệu cần nhập.

- Số lượng dữ liệu cần nhập.
- Mật độ dữ liệu.
- Yêu cầu về độ chính xác.
- Mục đích sử dụng dữ liệu.
- Thời gian cho phép đối với việc nhập dữ liệu.
- Kinh phí cho phép đối với việc nhập dữ liệu.



# CHƯƠNG 4

---

## Quản lý dữ liệu

### 4.1. Nguyên tắc quản lý dữ liệu

Quản lý dữ liệu là một chức năng quan trọng của tất cả các hệ thống thông tin trong đó có cả hệ thống thông tin địa lý. Nó giúp cho việc đưa dữ liệu vào và lấy dữ liệu ra từ hệ thống được tiến hành một cách thuận lợi, an toàn và hiệu quả. Về mặt chuyên môn, thuật ngữ “quản lý dữ liệu” bao gồm việc tổ chức, sắp xếp, tìm kiếm và bảo trì dữ liệu; thuật ngữ này còn bao hàm cả việc đảm bảo các thiết bị phần cứng, phần mềm và kiểm soát việc sử dụng các thiết bị này.

Chức năng quản lý dữ liệu của GIS được trợ giúp bởi hệ quản trị cơ sở dữ liệu; đó là một phần mềm cho phép một hoặc nhiều người làm việc với dữ liệu một cách có hiệu quả. Các hợp phần căn bản của hệ phải cung cấp các phương tiện để xác định nội dung của cơ sở dữ liệu, đưa vào dữ liệu mới, xóa dữ liệu cũ, hỏi về nội dung cơ sở dữ liệu và thay đổi nội dung cơ sở dữ liệu.

Bước đầu tiên trong quản lý dữ liệu là việc xác định nội dung của cơ sở dữ liệu với các thông tin cần thiết về khuôn dạng dữ liệu, nội dung dữ liệu và các hạn chế giá trị.

- Định nghĩa khuôn dạng dữ liệu đề cập đến kiểu dữ liệu như kiểu số, số nguyên, ký tự, số thập phân, ngày tháng,... và lượng bộ nhớ cần lưu trữ hay trình bày dữ liệu.
- Định nghĩa nội dung dữ liệu đề cập đến tên các trường hay các mục trong cơ sở dữ liệu. Nên sử dụng các tên cụ thể như: kinh độ, vĩ độ, độ pH, nồng độ DO,...
- Các hạn chế giá trị đề cập đến việc người sử dụng đưa vào trong hệ thống các hạn chế về giá trị dữ liệu để kiểm chứng các giá trị mới đưa vào. Ví dụ như các hạn chế về số tháng trong năm hay số phút trong một giờ hay kinh độ, vĩ độ của một vùng lãnh thổ.

Hệ quản trị cơ sở dữ liệu phải cung cấp các công cụ để đảm bảo sự an toàn, toàn vẹn, đồng bộ hóa, độc lập và giảm thiểu dư thừa dữ liệu.

- An toàn dữ liệu đề cập đến việc hạn chế các hình thức tiếp cận cơ sở dữ liệu bởi người sử dụng dữ liệu, bảo vệ dữ liệu khỏi bị tiết lộ một cách tình cờ hay cố ý và bị thay đổi hay phá hoại bởi những người không được ủy quyền. Như vậy, chỉ có những người có đầy đủ hiểu biết, thẩm quyền mới được phép thay đổi nội dung cơ sở dữ liệu.

- Để đảm bảo sự toàn vẹn của cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu kiểm tra các yếu tố được đưa vào để thực thi các ràng buộc cấu trúc cần thiết của dữ liệu bên trong như kiểm tra các giá trị cho phép trong các trường dữ liệu, ngăn ngừa việc xóa nút khi vẫn còn các cung được xác định dựa trên nút đó,...
- Sự đồng bộ hóa đề cập đến các hình thức bảo vệ chống lại sự không nhất quán có thể phát sinh do nhiều người sử dụng đồng thời cơ sở dữ liệu. Một ví dụ minh họa là khi có hai người cùng tiếp cận một lớp dữ liệu về sử dụng đất trong cơ sở dữ liệu, trong đó người thứ nhất thì cập nhật dữ liệu con người thứ hai thì tìm cách phân tích dữ liệu và như vậy, kết quả phân tích sẽ thay đổi theo thời gian và làm thất vọng người thứ hai nếu như không có một cơ chế cảnh báo hay ngăn ngừa người thứ hai tiếp cận cơ sở dữ liệu cho đến khi người thứ nhất hoàn tất thao tác của mình.
- Sự độc lập dữ liệu vật lý thể hiện ở chỗ sự lưu trữ dữ liệu và phần cứng điều khiển không được ảnh hưởng đến người sử dụng cơ sở dữ liệu. Nó cho phép ta thay đổi phần cứng khi cần thiết và khi công nghệ thay đổi thì không hề phải viết lại phần mềm điều khiển dữ liệu liên quan.
- Dư thừa dữ liệu là một điều hoàn toàn không mong muốn trong một cơ sở dữ liệu. Một ví dụ là trường hợp lưu dữ liệu vector spaghetti biểu diễn các đối tượng vùng. Sự dư thừa dữ liệu sẽ làm phức tạp việc cập nhật dữ liệu và giảm tốc độ xử lý. Do vậy, tối thiểu hóa sự dư thừa dữ liệu là một mục tiêu đặt ra đối với một cơ sở dữ liệu.

Trong quản lý cơ sở dữ liệu GIS, cần chú ý đến các vấn đề sau:

- Các dữ liệu bản đồ được quản lý theo hệ thống chồng ghép lớp và dễ dàng truy cập từ bất kỳ các lớp logic nào cùng với các topology của chúng;
- Các phương pháp truy cập dữ liệu gắn liền với hệ thống tổ chức quản lý chúng;
- Cơ sở dữ liệu tập trung và cơ sở dữ liệu phân tán của GIS;
- Các chuẩn dữ liệu và bảo trì, nâng cấp;
- Sự liên kết giữa các thành phần trong cơ sở dữ liệu và với các hệ thống khác;
- Bảo mật thông tin và quyền truy cập vào hệ thống cơ sở dữ liệu;
- Tính đa nhiệm của cơ sở dữ liệu trong GIS.

### ➤ Các yếu tố ảnh hưởng đến việc quản lý dữ liệu

Hiệu quả của việc lưu trữ, truy cập, xóa, sao chép hay cập nhật dữ liệu phụ thuộc vào nhiều yếu tố trong đó có thể kể đến hai yếu tố chính sau:

- Môi trường lưu trữ dữ liệu,
- Cấu trúc dữ liệu.

Cho đến nay, trong công nghệ thông tin nói chung và công nghệ GIS nói riêng, dữ liệu số thường được lưu trữ trong các môi trường sau:

#### a) Đĩa từ

Dữ liệu trên đĩa từ được tổ chức thành các rãnh và các khu vực. Các đầu từ dùng để đọc và viết dữ liệu lên đĩa. Tốc độ chuyển dữ liệu đến và từ đĩa đi phụ thuộc vào tốc độ quay của đĩa và mật độ dữ liệu dọc theo các rãnh.

Trong hệ thống lưu trữ dữ liệu bằng đĩa từ, sự tiếp cận dữ liệu là bất kỳ vì thế có thể điều khiển đầu từ tới một rãnh dữ liệu bất kỳ từ một rãnh bất kỳ khác.

Đĩa từ có thể là đĩa cứng hoặc đĩa mềm. Đĩa cứng được dùng để lưu trữ dữ liệu và chương trình cần cho việc tiếp cận trước mắt. Các đặc tính kỹ thuật của đĩa cứng là dung lượng nhớ tính bằng MB hay GB, tốc độ truy cập đo bằng mili giây (số mili giây càng thấp thì tốc độ càng nhanh), tốc độ truyền dữ liệu được đo bằng MB trên giây. Đĩa mềm được dùng để lưu trữ và trao đổi dữ liệu. Hiện nay, dung lượng đĩa mềm thông thường là 1,44 MB;

có một số loại là 2,88 MB và cao hơn. Tốc độ truy cập dữ liệu của đĩa mềm không cao như đĩa cứng. Ngoài ra, hiện nay trên thị trường còn có một số loại đĩa cứng di động khá phổ biến có khả năng giao tiếp với máy tính qua cổng USB.

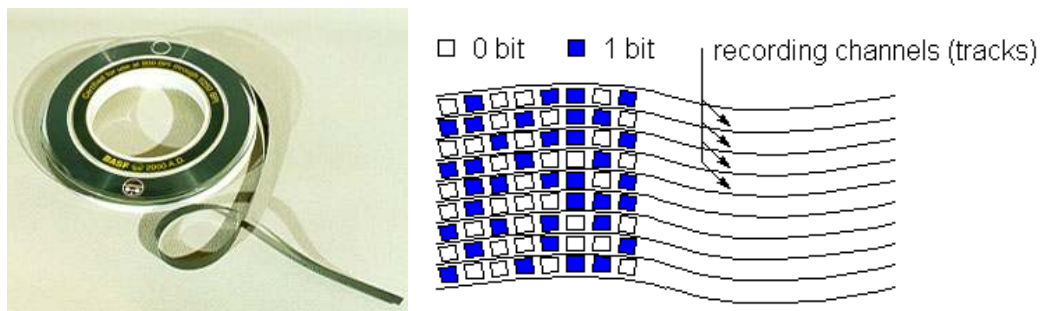


**Hình 4.1:** Một số loại đĩa từ thông dụng

#### b) Băng từ

Băng từ là một môi trường lưu trữ dữ liệu trong đó có thể truy cập dữ liệu theo cách tuần tự vì hệ thống lưu trữ phải đi qua hết chiều dài băng để định vị các yếu tố dữ liệu. Mật độ dữ liệu trên băng từ theo chuẩn công nghiệp là 1600-6250 bits/inch với 9 rãnh dữ liệu song song trên băng từ rộng 0,5 inch. 8 rãnh được dùng để lưu trữ dữ liệu và rãnh cuối cùng dùng để chống lỗi. Tốc độ đọc và ghi phổ biến là 25-125 inch/giây.

Băng từ chủ yếu được dùng để sao lưu cơ sở dữ liệu, lưu trữ lâu dài và chuyển dữ liệu giữa các hệ thống.



*Hình 4.2: Băng từ*

### c) Đĩa CD

Các đĩa này có thể lưu trữ hàng trăm MB dữ liệu. Để đọc được các đĩa CD cần phải có ổ đọc riêng còn muốn lưu trữ dữ liệu vào loại đĩa này thì phải có ổ ghi riêng. Đĩa CD rất thích hợp cho việc lưu trữ dữ liệu viễn thám và GIS.



Hiệu quả tìm kiếm dữ liệu nói chung phụ thuộc vào các yếu tố như khối lượng dữ liệu lưu trữ, phương pháp mã hóa dữ liệu, cấu trúc cơ sở dữ liệu và tính phức tạp của yêu cầu đặt ra.

- Khối lượng dữ liệu lưu trữ thể hiện ở số lượng các file và kích cỡ file dữ liệu trong cơ sở dữ liệu. Nó ảnh hưởng đến tốc độ tìm kiếm dữ liệu, đặc biệt là khi cần tìm kiếm hết cả cơ sở dữ liệu.
- Phương pháp mã hóa dữ liệu bao gồm quyết định về các loại biến cần lưu trữ cũng như cách thức lưu trữ các giá trị.
- Cấu trúc cơ sở dữ liệu liên quan đến các mô hình dữ liệu và cách thức tổ chức các file dữ liệu trong cơ sở dữ liệu địa lý (xem thêm phần cấu trúc dữ liệu và cơ sở dữ liệu) ảnh hưởng đến khối lượng dữ liệu và tốc độ tìm kiếm dữ liệu.
- Tính phức tạp của yêu cầu về cơ sở dữ liệu thể hiện ở loại và lượng các yêu cầu đặt ra như tìm kiếm một đối tượng, một tập hợp các đối tượng hay toàn bộ các đối tượng thỏa mãn các điều kiện nào đó.

Nhìn chung, một số bộ chương trình được phát triển tốt để tìm kiếm khá hiệu quả các dữ liệu phi không gian. Tuy nhiên, công việc tìm kiếm các đối tượng không gian hay các bộ đối tượng phức tạp hơn nhiều và việc tối ưu hóa tính năng tìm kiếm của hệ thống trong những điều kiện đó là một lĩnh vực nghiên cứu của hoạt động GIS.

## 4.2. Quản lý dữ liệu không gian

Dữ liệu không gian dưới dạng vector hay raster là dữ liệu về vị trí của các đối tượng địa lý như điểm, đường, vùng và được nhập vào GIS thông qua các con đường và bằng các công nghệ khác nhau. Việc tạo ra một cơ sở dữ liệu không gian trong GIS là một quá trình rất tốn kém về thời gian và công sức do vậy nhất thiết phải có các biện pháp, các công cụ mạnh để quản lý nó một cách hữu hiệu, đảm bảo hoạt động có kết quả của GIS.

Trong GIS, các dữ liệu được quản lý theo các lớp chuyên đề. Trong mỗi lớp chuyên đề, chỉ lưu trữ các thông tin về một chuyên đề nhất định như đất, địa giới, sông, đường giao thông,...

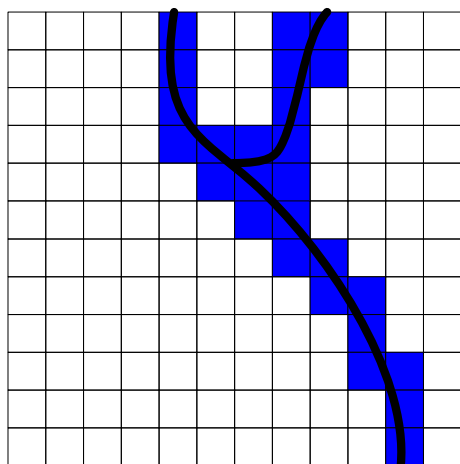
Các lớp dữ liệu như đất, nước, giao thông,... được tập hợp lại thành các lớp phủ và được tổ chức như một mảnh bản đồ. Các lớp thông tin trên còn là các lớp thông tin chuyên đề; trên một vùng lãnh thổ, chúng được thể hiện như các lớp phủ theo phương thẳng đứng. Do vậy, cần thiết phải có một hệ tọa độ để có thể chồng xếp các lớp thông tin đó lên nhau, phủ vùng lãnh thổ chúng thể hiện.

### ➤ *Sự chuyển đổi dữ liệu giữa các mô hình Raster và Vector*

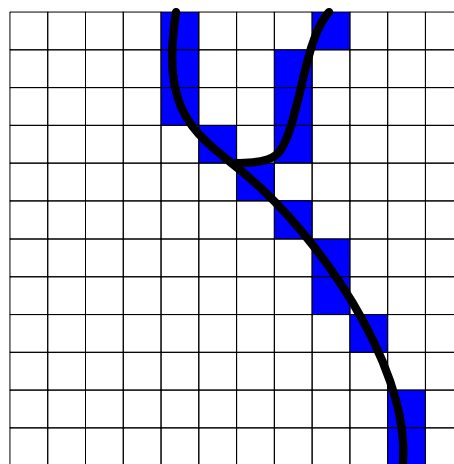
Dữ liệu không gian có thể được biến đổi qua lại giữa mô hình dữ liệu Raster và Vector. Sự chuyển đổi này là cần thiết trong việc quản lý dữ liệu không gian GIS khi chúng được lưu trữ trong những mô hình dữ liệu khác nhau. Sự chuyển đổi dữ liệu từ vector qua raster bao gồm việc xác định giá trị pixel cho mỗi vị trí liên quan đến các đối tượng vector. Đối tượng điểm trong mô hình vector có đặc trưng là không có kích thước trong khi đó các đối tượng điểm trong mô hình dữ liệu raster phải được xác định bởi một giá trị trong một ô raster (pixel); do đó các điểm ít nhất cũng phải có kích thước của các ô raster sau khi được chuyển đổi từ mô hình vector sang mô hình raster. Các đối tượng điểm thường được xác định bởi các pixel bao gồm tọa độ của điểm. Pixel mà đối tượng điểm xuất hiện trong nó được gán bởi một con số hoặc một mã để nhận biết thuộc tính của điểm xuất hiện tại vị trí pixel đó. Nếu kích thước ô quá lớn thì hai hay nhiều đối tượng điểm vector có thể sẽ rơi vào cùng một pixel hay là một công cụ nhận biết pixel nào đó được sử dụng hoặc một sự sắp xếp phức tạp hơn theo kiểu phân chia và đánh số được thực hiện. Thông thường, kích thước của một ô được lựa chọn sao cho kích thước của đường chéo ô nhỏ hơn so với khoảng cách giữa hai đối tượng điểm gần nhất.

Các đối tượng đường vector trong một lớp dữ liệu cũng có thể được chuyển đổi thành một mô hình dữ liệu raster. Các ô raster có thể được mã hóa bằng nhiều tiêu chí khác nhau. Một phương pháp đơn giản là xác định một giá trị cho một pixel nếu đường vector cắt bất kỳ một phần nào của pixel đó (Hình 4.3a). Phương pháp này đảm bảo duy trì các đối tượng đường được kết nối liên tục trong khuôn dạng dữ liệu raster. Quy tắc chuyển đổi này thường dẫn đến việc tạo ra các đối tượng đường rộng hơn so với cần thiết bởi vì một vài ô gần kề nhau có thể được xác định như là một phần của đối tượng đường; đặc biệt là khi đối tượng đường

ngoài ra, quanh co gần các cạnh của các ô raster. Một quy tắc chuyển đổi khác có thể được ứng dụng là chỉ xác định một ô có liên quan đến đối tượng đường khi tâm của ô đó gần với đoạn của đối tượng đường đi qua (Hình 4.3b). “Gần” có thể được định nghĩa như là khoảng cách của một số ô phụ; chẳng hạn như 1/3 chiều rộng của một ô. Các đường chỉ đi cắt qua góc của một ô sẽ không được ghi lại như là một pixel. Điều này có thể dẫn đến việc thu được bộ dữ liệu raster có các đối tượng đường mảnh hơn và thường thì kết quả mang lại là các đối tượng đường không liên tục.



Hình 4.3a



Hình 4.3b

**Hình 4.3:** Chuyển đổi dữ liệu vector sang raster

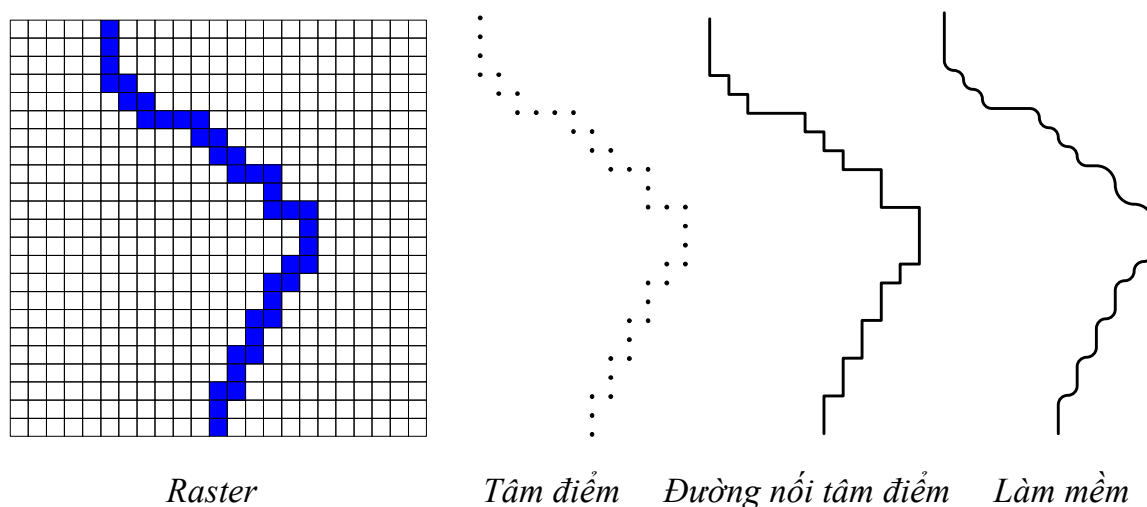
Kết quả đầu ra của phép chuyển đổi vector sang raster phụ thuộc vào thuật toán đầu vào được sử dụng. Ta có thể thu được các lớp dữ liệu đầu ra khác nhau khi sử dụng các thuật toán chuyển đổi khác nhau mặc dù là dữ liệu đầu vào mà ta sử dụng là hoàn toàn giống nhau. Đây là một điểm quan trọng và cần phải ghi nhớ khi ứng dụng bất kỳ một phép toán không gian nào. Những sự khác nhau rất nhỏ trong thuật toán hoặc là các giới hạn được định nghĩa có thể dẫn đến các kết quả khá sai khác với nhau. Những thay đổi nhỏ trong khoảng cách hay quy tắc xác định trong quá trình chuyển đổi vector sang raster có thể mang lại những khác biệt lớn trong các bộ dữ liệu đầu ra, mặc dù là cùng dữ liệu đầu vào. Thường thì không có một ưu thế rõ ràng nào để xác định phương pháp tốt nhất. Các thử nghiệm dựa vào kinh nghiệm cũng như các kinh nghiệm có được từ trước chính là sự hướng dẫn hữu ích cho việc xác định phương pháp tối ưu đối với một bộ dữ liệu cho trước hay một vấn đề cụ thể trong quá trình chuyển đổi. Chính sự thoải mái và không bị ràng buộc bởi một quy tắc cụ thể nào của thao tác không gian trong GIS đã mang lại các công cụ mạnh và dễ sử dụng cho người dùng. Người sử dụng GIS phải luôn ghi nhớ rằng các công cụ đó có thể trở nên hiệu quả hơn trong việc cung cấp các kết quả chính xác nhưng cũng có thể góp phần tạo ra các lỗi trong khi sử dụng.

Các đối tượng vùng được chuyển đổi từ vector sang raster theo phương pháp tương tự sử dụng cho đối tượng đường vector. Đường viền giữa các vùng khác nhau được xác định bằng cách sử dụng phương pháp được mô tả ở hình 4.3a và hình 4.3b. Sau đó, vùng bên trong được nhận dạng và mỗi ô thuộc vùng bên trong

được xác định bởi một giá trị nào đó. Lưu ý rằng các ô đường viền bao gồm đường biên giữa các vùng phải được xác định. Tương tự như sự chuyển đổi vector sang raster của các đối tượng đường, có một số phương pháp để xác định một ô raster thuộc đối tượng vùng hay không. Một phương pháp phổ biến xác định một ô thuộc vùng nếu hơn một nửa ô phân bố trong vùng vector. Một phương pháp phổ biến khác xác định một ô thuộc một đối tượng vùng nếu bất kỳ một phần nào của ô đó phân bố trong vùng thuộc đối tượng vùng vector. Và các kết quả chuyển đổi thu được sẽ biến đổi tùy theo phương pháp được sử dụng.

Ngoài ra, dữ liệu cũng được chuyển đổi theo cách ngược lại, có nghĩa là dữ liệu raster có thể được chuyển đổi thành dữ liệu vector. Các đối tượng điểm, đường hay vùng được mô tả bằng các ô raster được chuyển đổi thành các cấu trúc và tọa độ dữ liệu vector tương ứng. Đối tượng điểm được mô tả như là một ô raster. Mỗi đối tượng điểm vector thường được xác định bởi một tọa độ của tâm điểm của pixel tương ứng.

Các đối tượng đường được mô tả trong môi trường raster có thể được chuyển đổi thành các đường vector. Sự chuyển đổi thành đường vector thường bao gồm việc xác nhận các ô lưới nối kết liên tục với nhau hình thành nên một đường. Tâm điểm của các ô thường được xác định như là vị trí hay các đỉnh dọc theo đường. Sau đó, các đường có thể được làm mềm bằng cách sử dụng một thuật toán để xóa bỏ hiệu ứng bậc cầu thang.



**Hình 4.4:** Chuyển đổi dữ liệu raster sang vector

### 4.3. Quản lý dữ liệu thuộc tính

Trong GIS, các dữ liệu thuộc tính thường được lưu trữ dưới dạng bảng do vậy thường được gọi là dữ liệu bảng. Trong các bảng thuộc tính, mỗi dòng trong bảng biểu diễn một đối tượng và mỗi cột là một thuộc tính.

Ví dụ sau là các bảng dữ liệu thuộc tính mô tả một số khu rừng, các đường mòn trong khu rừng, các loại hình giải trí có thể tìm thấy trên các đường mòn

đó,... Những dữ liệu này có thể được xây dựng một cách vật lý để trả lời các câu hỏi liên quan thường xuyên đặt ra như “đường mòn nào bị cắt ngang bởi thác nước?”.

### ***Rừng***

<b>Tên rừng</b>	<b>Vị trí</b>	<b>Diện tích</b>
Nantahala	Bắc Carolina	184.447
Cherokee	Bắc Carolina	92.271

### ***Đặc điểm đường mòn***

<b>Tên đường mòn</b>	<b>Độ khó</b>	<b>Rừng</b>	<b>Đặc trưng</b>
Bryson's Knob	Dễ, Vừa	Nantahala	PC, RG
Slickrock Falls	Vừa	Cherokee	TN, RG
North Frok	Vừa	Nantahala	-
Cade's Cove	Dễ	Cherokee Nantahala	RG, HD
Appalachian	Vừa, Khó	Nantahala Cherokee	TN, RG, PC, HD, CT

### ***Đặc trưng giải trí***

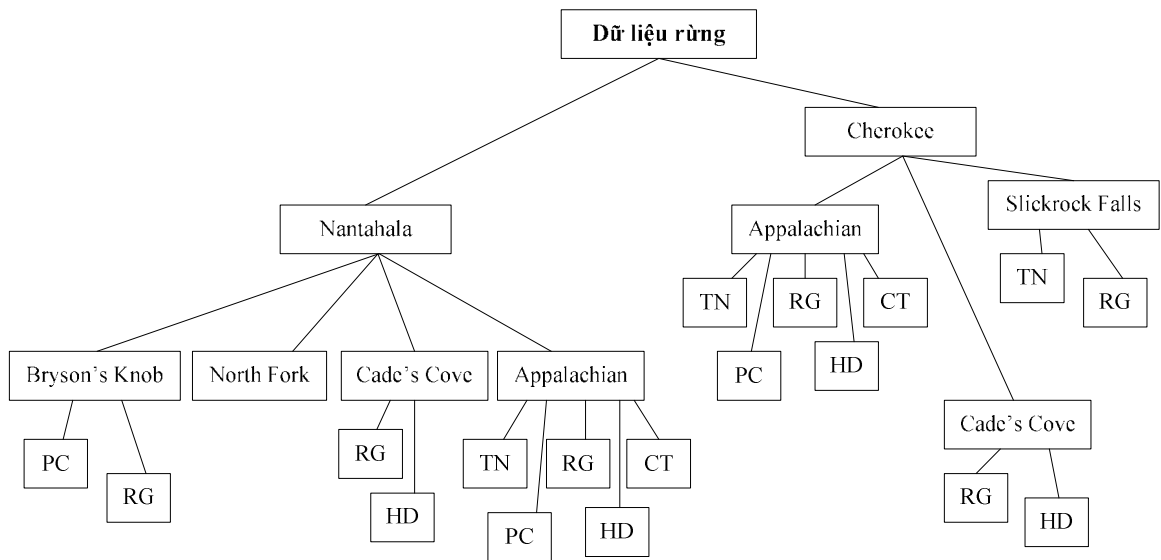
<b>Đặc trưng</b>	<b>Mô tả</b>	<b>Hoạt động</b>
TN	Thác nước	Chụp ảnh, bơi lội
RG	Rừng già	Chụp ảnh, đi bộ
PC	Phong cảnh từ trên cao	Chụp ảnh, ngắm cảnh
HD	Thế giới hoang dã	Chụp ảnh, ngắm chim
CT	Cắm trại	Cắm trại

### **➤ *Dữ liệu thuộc tính có thể quản lý bằng các hệ cơ sở dữ liệu sau:***

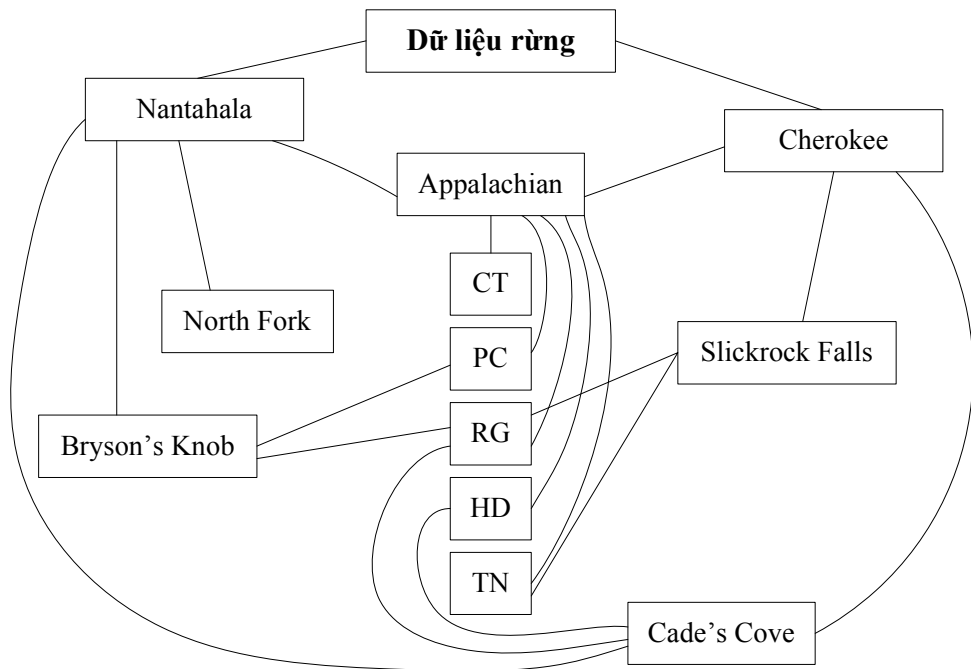
- *Hệ quản trị cơ sở dữ liệu phân cấp và mạng*

Dữ liệu được truy cập theo một cấu trúc phân cấp bằng cách đi từ trên xuống theo các nhánh của sự phân cấp. Theo cách đó, các dữ liệu con của một dữ liệu mẹ được truy cập một cách nhanh chóng. Trong ví dụ trên, tất cả các đường mòn trong rừng Cherokee đều liên kết trực tiếp với thực thể mẹ (Hình 4.5) và có thể được truy cập trực tiếp từ nhánh đó của cây phân cấp. Tuy nhiên, ví dụ đơn giản này cũng cho ta thấy một số hạn chế của kiểu cơ sở dữ liệu phân cấp này như các thông tin thuộc tính được ghi lại nhiều lần trong cơ sở dữ liệu, gây lãng phí bộ nhớ và phức tạp trong việc chỉnh sửa và cập nhật.





**Hình 4.5:** Dữ liệu rừng trong hệ cơ sở dữ liệu phân cấp



**Hình 4.6:** Dữ liệu rừng trong hệ cơ sở dữ liệu mạng

Kiểu cơ sở dữ liệu mạng khác với kiểu phân cấp; trong đó, phần lớn dữ liệu con có thể có nhiều hơn một dữ liệu mẹ (Hình 4.6). Mỗi thực thể có thể là một nút trong mạng lưới và được liên kết với các thực thể thích hợp khác trong toàn mạng. Kiểu cơ sở dữ liệu mạng có ít dữ liệu dư thừa hơn so với kiểu phân cấp nhờ vào ưu điểm của các mối quan hệ thêm vào; và nó chỉ đòi hỏi một bản sao của một thực thể cụ thể. Trong ví dụ về dữ liệu rừng trên, đặc trưng giải trí RG (rừng già) được

mô tả chỉ một lần và được kết nối đến các đường mòn tương ứng trong cơ sở dữ liệu.

- *Hệ quản trị cơ sở dữ liệu quan hệ*

Một cơ sở dữ liệu quan hệ mô tả dữ liệu trong các bản và được biết đến như là các quan hệ. Mỗi thực thể được mô tả bởi một hàng và các thuộc tính kèm theo trên các cột tương ứng. Trong ví dụ về cơ sở dữ liệu rừng, có thể có các bản dữ liệu như dưới đây. Các bản có quan hệ với nhau thông qua các khóa (có thể là một hay nhiều trường) và được sử dụng để nhận biết các bản ghi.

### ***Rừng***

<b>Tên rừng</b>	<b>ID rừng</b>	<b>Vị trí</b>	<b>Diện tích</b>
Nantahala	1	Bắc Carolina	184.447
Cherokee	2	Bắc Carolina	92.271

### ***Đường mòn***

<b>Tên đường mòn</b>	<b>ID rừng</b>
Bryson's Knob	1
Slickrock Falls	2
North Frok	1
Cade's Cove	1
Cade's Cove	2
Appalachian	1
Appalachian	2

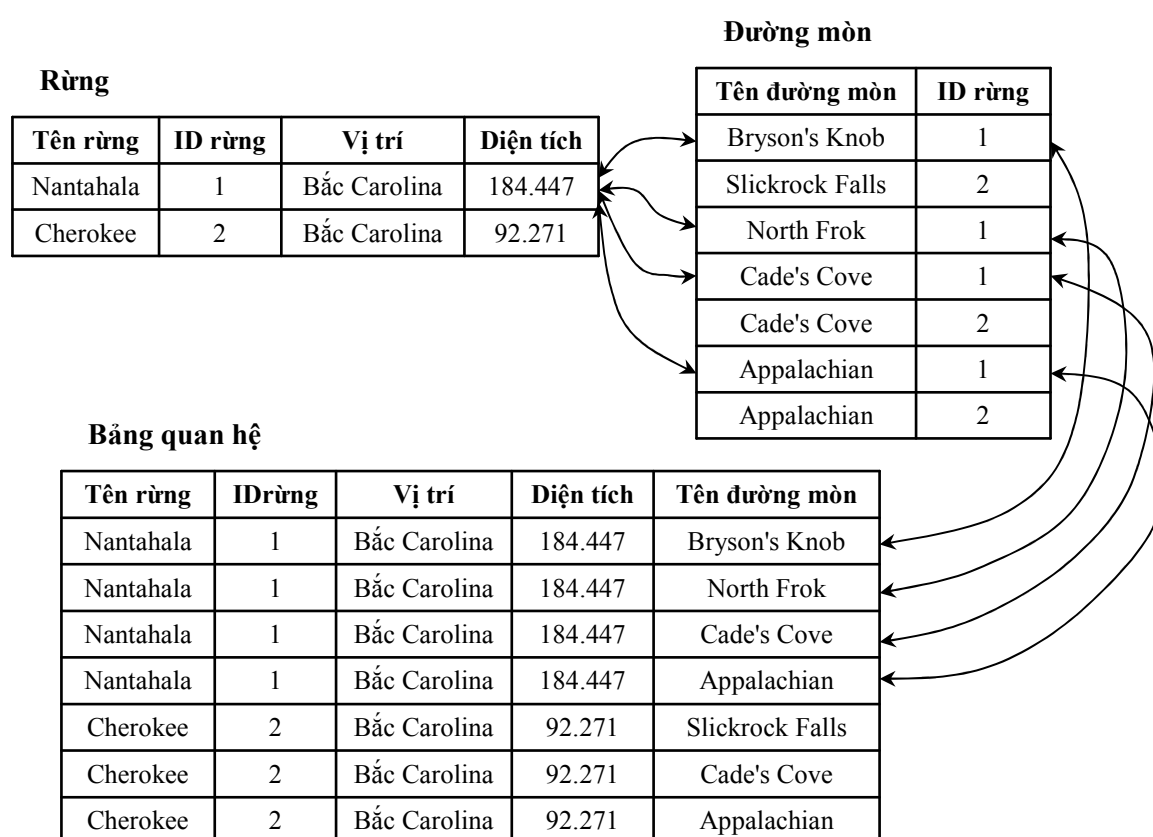
### ***Đặc điểm***

<b>Tên đường mòn</b>	<b>Đặc trưng</b>	<b>Độ khó</b>
Bryson's Knob	PC	Dễ, Vừa
Bryson's Knob	RG	Dễ, Vừa
Slickrock Falls	RG	Vừa
Slickrock Falls	TN	Vừa
North Frok	-	Vừa
Cade's Cove	RG	Dễ
Cade's Cove	HD	Dễ
Appalachian	TN	Vừa, Khó
Appalachian	RG	Vừa, Khó
Appalachian	PC	Vừa, Khó
Appalachian	HD	Vừa, Khó
Appalachian	CT	Vừa, Khó

### Đặc trưng giải trí

Đặc trưng	Mô tả	Hoạt động 1	Hoạt động 2
TN	Thác nước	Chụp ảnh	Bơi lội
RG	Rừng già	Chụp ảnh	Đi bộ
PC	Phong cảnh từ trên cao	Chụp ảnh	Ngắm cảnh
HD	Thế giới hoang dã	Chụp ảnh	Ngắm chim
CT	Cắm trại	Cắm trại	-

Hình dưới đây mô tả các khu rừng và đường mòn trong cấu trúc của một bảng quan hệ với khóa chính là **ID rừng** trong bảng **Rừng**. Cần lưu ý rằng không phải tất cả dữ liệu phải được mô tả trong một bảng và cần thiết phải có một kỹ thuật để liên kết giữa các bảng với nhau.



**Hình 4.7:** Dữ liệu rừng và đường mòn trong cấu trúc bảng quan hệ

Các dữ liệu địa lý, kể cả dữ liệu không gian và thuộc tính rất dễ bị mất do lỗi về thiết bị, môi trường và chương trình máy tính, cũng như sai sót của người sử dụng. Do đó, cần có biện pháp an toàn chống lại các sai sót đó; cụ thể là cần sao lưu dữ liệu thường kỳ các file dữ liệu, hàng ngày sao lưu các thay đổi cơ sở dữ liệu từ các đĩa lên băng, hàng tuần sao lưu toàn bộ cơ sở dữ liệu. Các đĩa sao lưu cần để ở những có điều kiện thích hợp, những nơi chịu lửa và không để cùng với dữ liệu gốc.

# CHƯƠNG 5

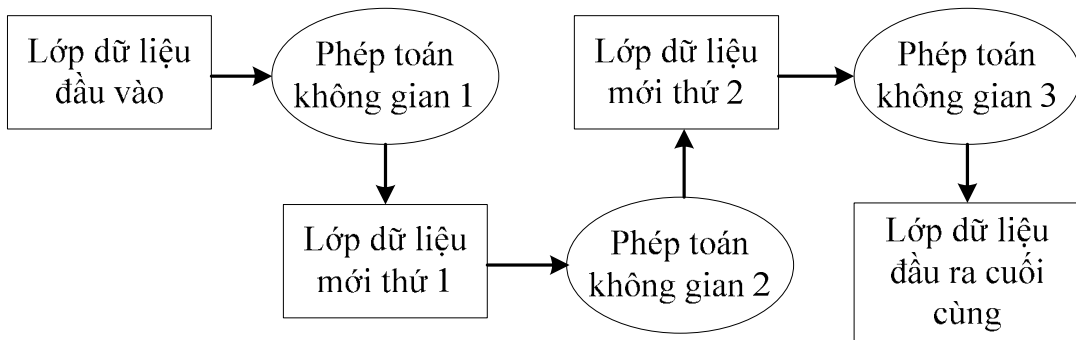
## Phân tích dữ liệu

### 5.1. Giới thiệu chung

Phân tích dữ liệu là rút ra những thông tin và kết luận từ việc nghiên cứu dữ liệu và là công việc của mỗi ngành nghề, mỗi nhà chuyên môn khi ứng dụng GIS. Phân tích dữ liệu và mô hình hóa không gian được xem là phần cốt lõi, phần trọng tâm và là chức năng quan trọng nhất của GIS.

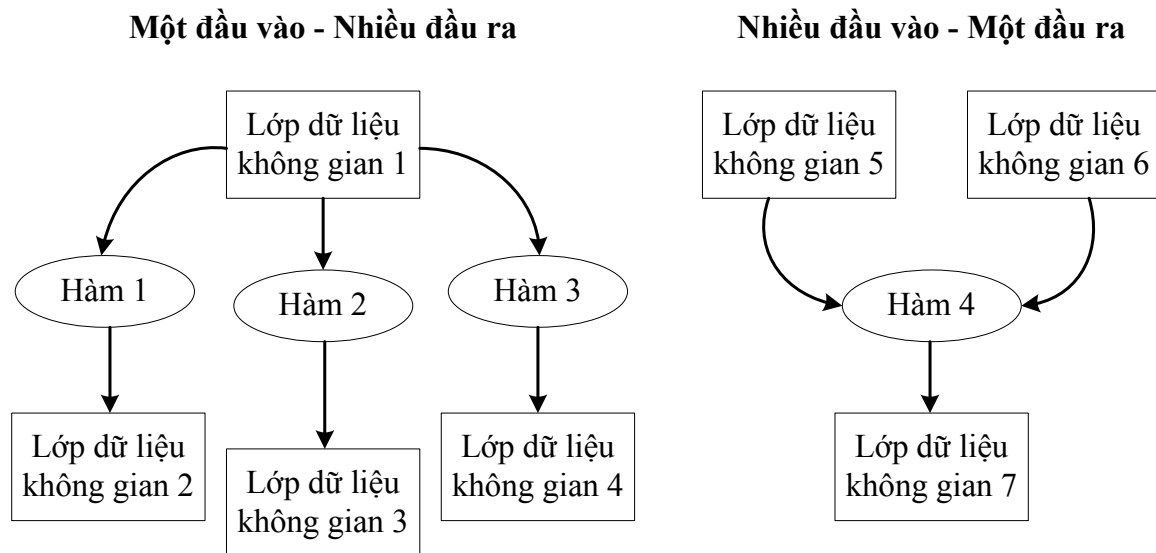
Phân tích dữ liệu không gian bao gồm việc sử dụng các phép toán để sắp xếp các dữ liệu đó cũng như các dữ liệu thuộc tính có liên quan. Đa số các phân tích không gian thường được ứng dụng để giải quyết các vấn đề cụ thể, ví dụ như nhận biết các vùng có tính bất ổn về an ninh cao, hay đưa ra danh sách các đoạn đường cần sửa chữa lại, hay là chọn một vị trí tốt nhất cho hoạt động kinh doanh.

Các phép toán không gian có thể được sử dụng liên tiếp nhau để giải quyết vấn đề nào đó. Mỗi phép toán không gian sẽ tạo ra một sản phẩm đầu ra và sản phẩm đó có thể được sử dụng như dữ liệu đầu vào của các phép toán khác. Một phần thách thức của việc phân tích dữ liệu không gian đó là việc lựa chọn các phép toán không gian thích hợp và ứng dụng chúng theo các trật tự thích hợp.



**Hình 5.1:** Chuỗi liên tục các phép toán không gian thường được sử dụng để thu được lớp dữ liệu cuối cùng mong muốn

Việc phân tích dữ liệu không gian thường sử dụng dữ liệu từ một hay nhiều lớp dữ liệu để tạo ra dữ liệu đầu ra. Phép phân tích có thể gồm có một phép toán đơn lẻ sử dụng cho một lớp dữ liệu hoặc nhiều phép toán có khả năng kết hợp dữ liệu đầu vào từ nhiều lớp dữ liệu để tạo ra dữ liệu đầu ra mong muốn.



**Hình 5.2:** Nhiều phép phân tích dữ liệu không gian có số lượng lớn các phép toán và hàm không gian

Như được minh họa trong hình 5.2, trong các phép phân tích dữ liệu không gian, có thể có một hay nhiều dữ liệu đầu vào hay đầu ra. Nhiều phép toán chỉ đòi hỏi một lớp dữ liệu đầu vào và tạo ra một lớp dữ liệu đầu ra; ví dụ như phép chuyển đổi dữ liệu vector sang raster. Cũng có các phép toán tạo ra một vài lớp dữ liệu đầu ra từ một đầu vào hay đòi hỏi một số đầu vào để tạo ra một lớp dữ liệu đầu ra. Các hàm phân tích địa hình có thể sử dụng lưới raster về độ cao như là một lớp dữ liệu đầu vào và tạo ra cả dữ liệu độ dốc (mỗi pixel có độ dốc như thế nào) và dữ liệu về hướng địa hình (độ dốc ở một vị trí nào đó có hướng như thế nào). Lớp dữ liệu trung bình là một ví dụ của việc sử dụng nhiều lớp dữ liệu đầu vào để tạo nên một lớp dữ liệu đầu ra.

Trong GIS, việc phân tích hay khai thác dữ liệu có thể được thực hiện ở các mức độ sau:

- Dữ liệu thuộc tính trong các bảng được sắp xếp lại để trình bày trong các báo cáo hay sử dụng ở các hệ máy tính khác.
- Các thao tác được thực hiện trên các dữ liệu hình học hay ở chế độ tìm kiếm hay vì các mục đích tính toán.
- Các thao tác logic, số học và thống kê được thực hiện ở các bảng thuộc tính.
- Hình học và bảng thuộc tính được dùng chung để lập các bộ dữ liệu mới dựa trên các thuộc tính gốc và phát sinh; hay lập bộ dữ liệu mới dựa trên các mối quan hệ địa lý.

Nói cách khác, phân tích dữ liệu bằng GIS có thể được xếp thành ba nhóm: hỏi đáp cơ sở dữ liệu, lập bản đồ phát sinh và mô hình hóa quá trình. Hỏi đáp cơ sở dữ liệu là đơn thuần lấy ra thông tin có sẵn trong cơ sở dữ liệu. Lập bản đồ phát

sinh là quá trình tạo ra các lớp dữ liệu mới từ các lớp dữ liệu cũ bằng cách lấy thông tin có sẵn và thêm vào đó thông tin mới là sự hiểu biết về mối quan hệ giữa các yếu tố cơ sở dữ liệu. Ví dụ như ta có thể tạo ra bản đồ nguy cơ xói mòn đất từ bản đồ đất, bản đồ độ cao địa hình và bản đồ chế độ lượng mưa nếu ta biết quan hệ giữa các nhân tố đó. Mô hình hóa quá trình là một lĩnh vực mới của GIS dựa trên khái niệm là cơ sở dữ liệu không chỉ biểu diễn môi trường thực mà chính nó cũng là một môi trường, do đó, cơ sở dữ liệu GIS đóng vai trò như một phòng thí nghiệm để tìm hiểu các quá trình.

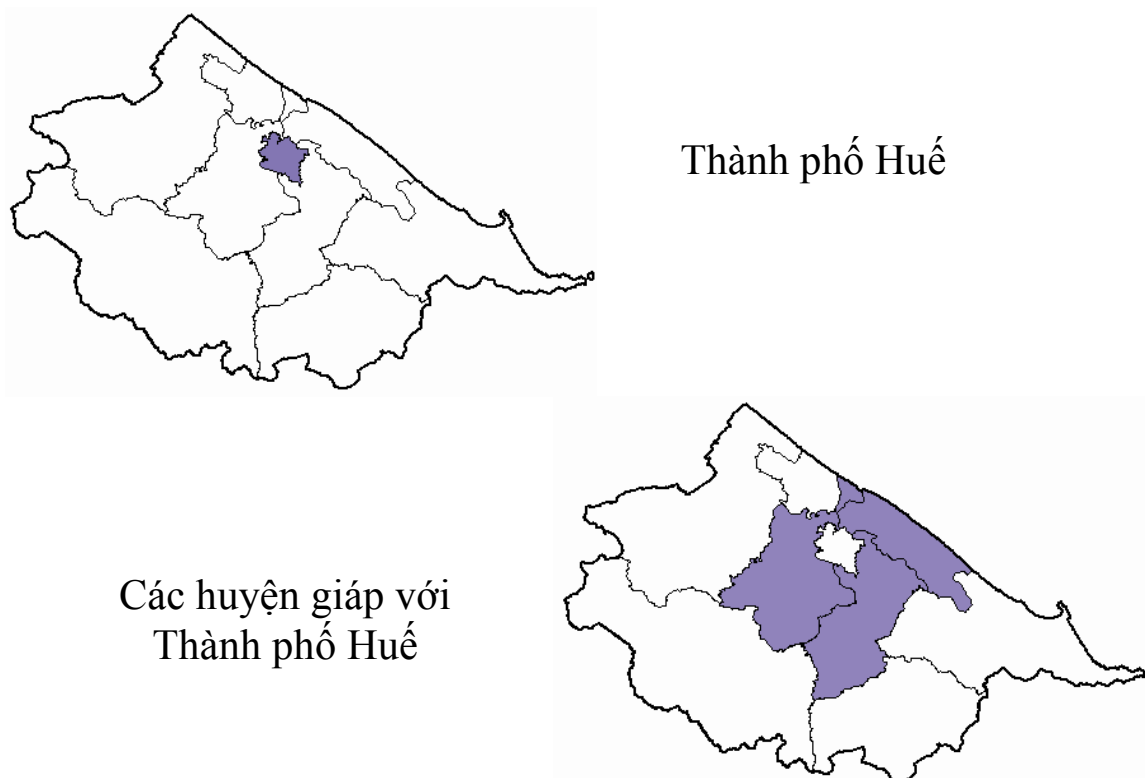
## 5.2. Các phép phân tích dữ liệu cơ bản

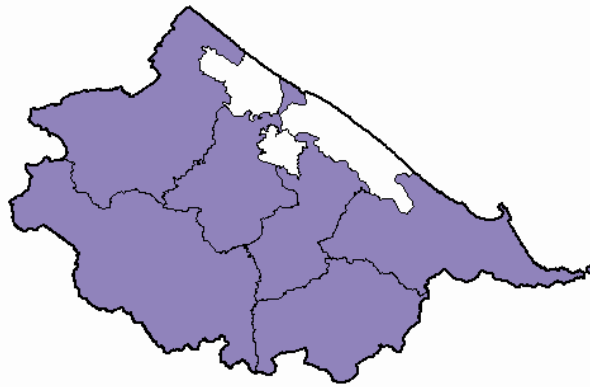
### 5.2.1. Lựa chọn và phân loại

#### a. Lựa chọn

Phép lựa chọn bao gồm việc nhận biết các thực thể thỏa mãn một hay nhiều điều kiện hay tiêu chí nào đó. Các thuộc tính hay đặc điểm hình học của các thực thể được kiểm tra dựa vào các tiêu chí và chỉ những gì thỏa mãn các tiêu chí mới được lựa chọn. Các thực thể được lựa chọn đó có thể được ghi lại lên trên một lớp dữ liệu mới hoặc dữ liệu hình học hay thuộc tính của chúng được lưu lại theo một vài cách khác nhau.

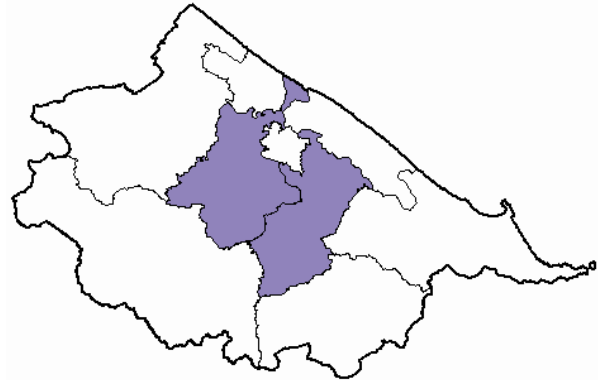
Hình 5.3 minh họa một ví dụ của phép lựa chọn bao gồm cả phần thuộc tính của một bộ dữ liệu không gian. Hai điều kiện được sử dụng và các đối tượng thỏa mãn cả hai điều kiện đó được lựa chọn.





Các huyện có diện tích lớn hơn 400 Km<sup>2</sup>

Các huyện giáp với Thành phố Huế và có diện tích lớn hơn 400 Km<sup>2</sup>



**Hình 5.3:** Một ví dụ về phép lựa chọn dựa trên nhiều điều kiện

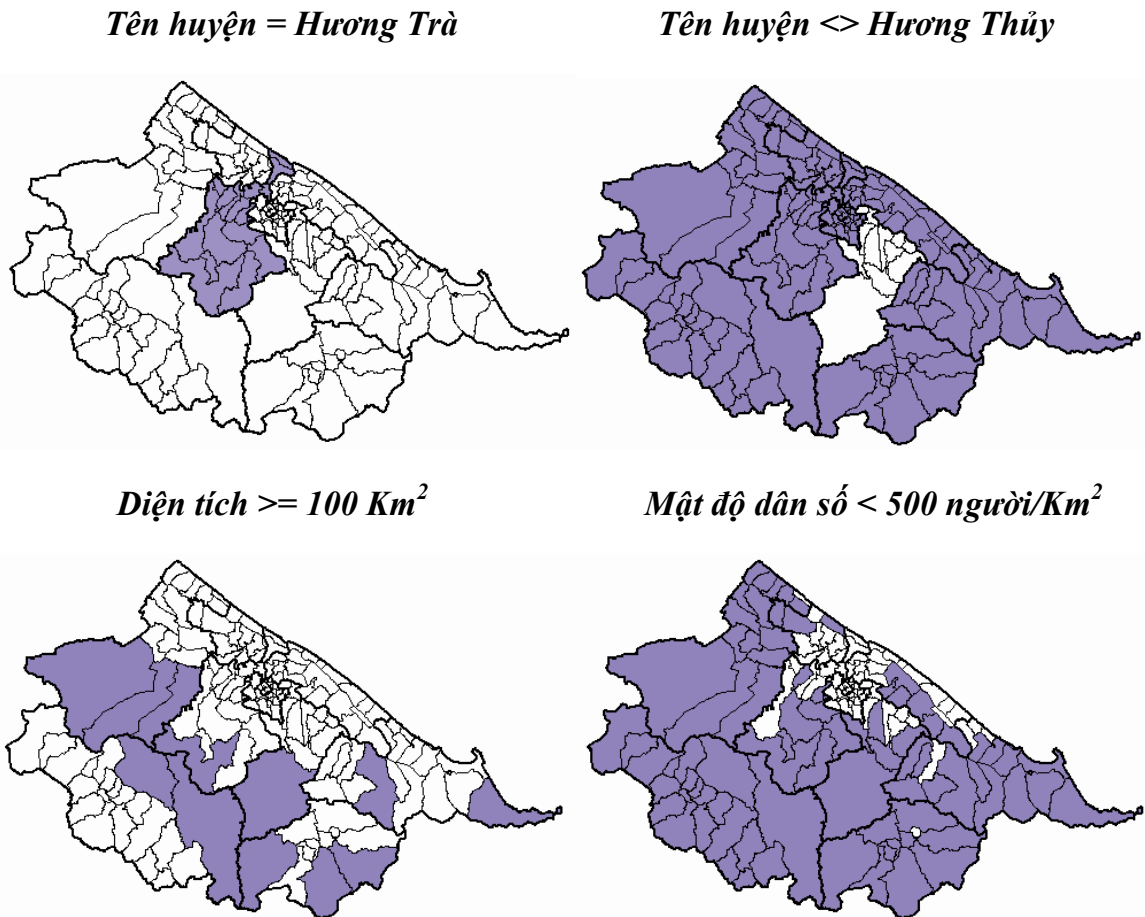
Hình thức đơn giản nhất của việc lựa chọn là phép truy vấn trực quan trên màn hình. Một lớp dữ liệu sẽ được hiển thị và các thuộc tính được lựa chọn thông qua một toán tử do người lập trình định nghĩa. Toán tử đó sử dụng một thiết bị trợ để định vị vị trí con trỏ chuột thông qua một đặc tính nào đó có liên quan (giả sử như động tác click chuột hay nhấn một phím nào đó trên bàn phím) và gửi đi một câu lệnh để tiến hành thao tác lựa chọn. Phép truy vấn trực quan trên màn hình thường dùng để thu thập các thông tin về các đối tượng cụ thể cũng như tương tác, cập nhật cho các dữ liệu thuộc tính hay không gian.

Các phép truy vấn cũng có thể được định rõ bởi các điều kiện một cách đơn lẻ đối với các thành phần phi không gian. Các phép lựa chọn này đa số thường dựa vào các bảng dữ liệu thuộc tính của một lớp hay nhiều lớp dữ liệu. Các thuộc tính của mỗi đối tượng được so sánh với nhóm các điều kiện, nếu thuộc tính thỏa mãn điều kiện thì chúng được lựa chọn, còn nếu chúng không thỏa điều kiện thì được xếp vào nhóm không được lựa chọn. Các dữ liệu được lựa chọn thường được sử dụng vào một hay nhiều mục đích nào đó và chúng được lưu trữ vào một file riêng biệt, được xóa và chỉnh sửa theo các cách khác nhau.

Thực tế thì phép lựa chọn bảng thuộc tính có quan hệ không gian với các đối tượng địa lý bởi vì mỗi bảng ghi trong bảng thuộc tính đều được liên kết với một đối tượng địa lý nào đó có liên quan. Lựa chọn một bảng ghi trong bảng thuộc tính cũng có thể được dùng để lựa chọn một hay nhiều pixel, điểm, đường hay vùng. Các phép lựa chọn không gian có thể được kết hợp với phép lựa chọn bảng thuộc tính để nhận biết nhóm các đối tượng địa lý được lựa chọn.

### ➤ Đại số tập hợp

Các điều kiện lựa chọn thường được sử dụng dưới hình thức đại số tập hợp. Đại số tập hợp sử dụng các toán tử: nhỏ hơn ( $<$ ), nhỏ hơn hoặc bằng ( $<=$ ), lớn hơn ( $>$ ), lớn hơn hoặc bằng ( $>=$ ), bằng ( $=$ ) và không bằng ( $<>$ ). Các toán tử này có thể được dùng riêng lẻ hay gộp chung để lựa chọn đối tượng.



**Hình 5.4:** Ví dụ về các biểu thức trong đại số tập hợp và kết quả đầu ra

Cần lưu ý rằng toán tử lớn hơn, lớn hơn hoặc bằng ( $>$ ,  $>=$ ) hay nhỏ hơn, nhỏ hơn hoặc bằng ( $<$ ,  $<=$ ) không sử dụng cho các dữ liệu địa danh và tên hay màu sắc,... Chỉ có các toán tử bằng ( $=$ ) và khác ( $<>$ ) được sử dụng cho các loại biến số đó. Tất cả các toán tử đại số tập hợp đều sử dụng được cho các dữ liệu về thứ tự, thời gian, tỷ lệ.

### ➤ Đại số logic

Đại số logic sử dụng điều kiện OR, AND, NOT và NOR để lựa chọn đối tượng. Các biểu thức logic thường được sử dụng để gộp nhiều điều kiện và tạo ra các lựa chọn không gian ghép. Biểu thức logic gồm có một bộ các toán tử logic, các biến và có thể có cả các giá trị vô hướng hay hằng số.



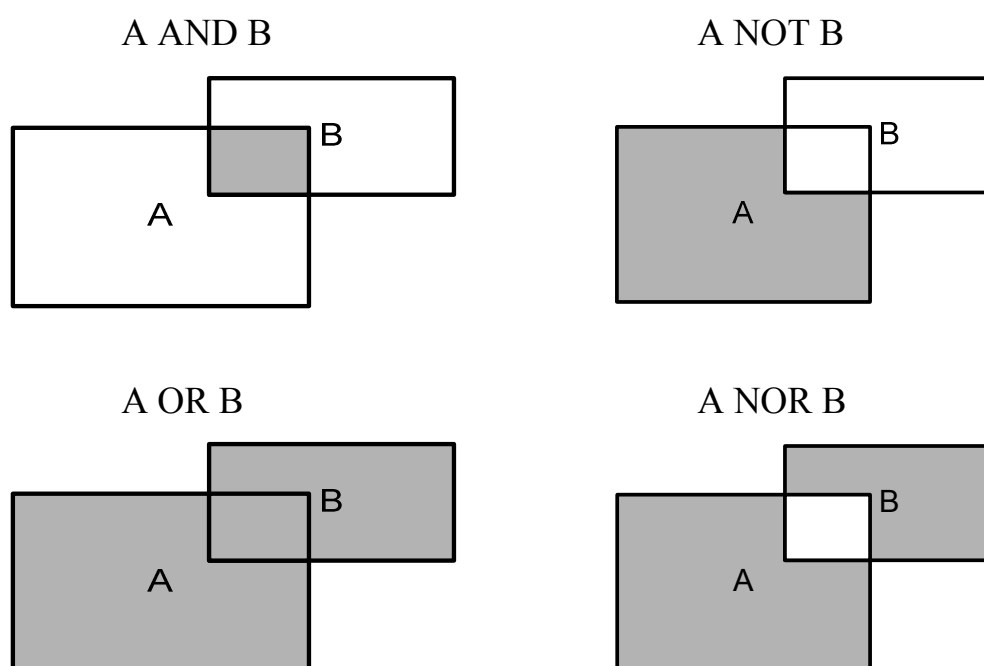
Các biểu thức logic được xác định bằng cách quy về giá trị đúng hoặc sai của biểu thức đối với từng điều kiện. Có thể lấy một số ví dụ cho biểu thức logic như sau:

(diện tích > 100.000) AND (giá trị nông sản < 10 tỉ)

hay NOT (huyện = Hương Trà)

hay [(lượng mưa > 1000) AND (thuế = thấp)] OR [(giá đất < 65.000) AND NOT (an ninh = thấp)]

Có thể dùng biểu đồ để minh họa cho các khai báo logic như sau:

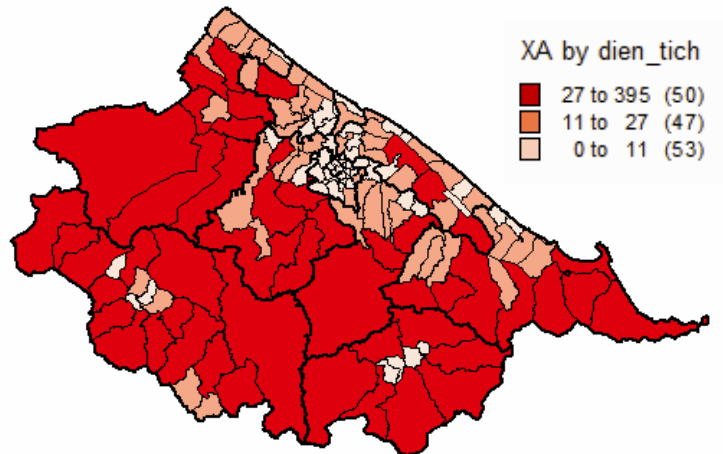


*Hình 5.5: Minh họa các khai báo logic*

### ***b. Phân loại***

Phân loại là một phép toán dữ liệu không gian thường được sử dụng kết hợp với phép lựa chọn. Phép phân loại còn được biết đến như là việc phân loại lại hay ghi lại với mục đích phân chia các đối tượng địa lý dựa vào một hay nhiều điều kiện nào đó.

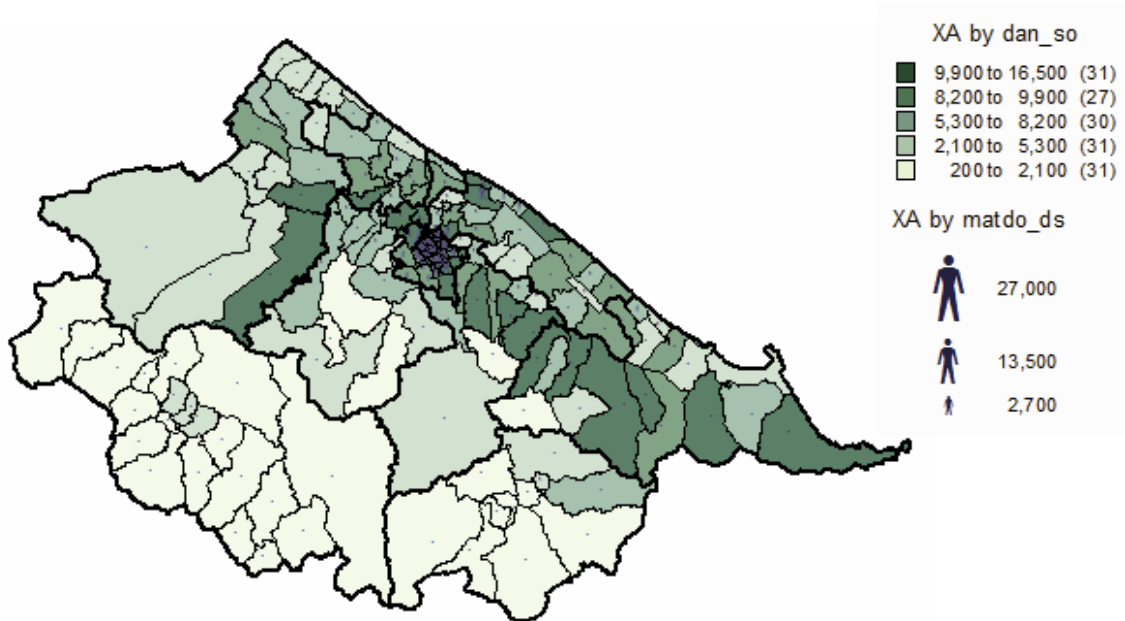
Ví dụ, tất cả các xã, phường có diện tích lớn hơn 27 Km<sup>2</sup> được phân loại vào nhóm có diện tích lớn, diện tích từ 11 Km<sup>2</sup> đến 27 Km<sup>2</sup> được xếp vào nhóm có diện tích trung bình và từ 0 đến 11 Km<sup>2</sup> được xếp vào nhóm có diện tích nhỏ. Như vậy, kết quả phân loại có thể như sau:



**Hình 5.6:** Kết quả phân loại phường xã theo diện tích đất

Phép phân loại có thể được sử dụng cho nhiều mục đích khác nhau. Một trong những mục đích phổ biến đó là nhóm gộp các đối tượng để hiển thị hay xuất bản các bản đồ. Các đối tượng này thường có một thuộc tính chung và mục đích cần đạt được là hiển thị chúng theo một màu sắc hay biểu tượng thống nhất, như vậy các đối tượng tương tự nhau sẽ được nhận biết theo nhóm. Màu sắc hay hoa văn được hiển thị thường phụ thuộc vào giá trị của một hay nhiều thuộc tính. Một dãy màu sắc biến chuyển đậm nhạt hay từ màu này sang màu khác sẽ được lựa chọn và các giá trị tương ứng cho mỗi thuộc tính cụ thể được xác định. Sau đó, bản đồ sẽ được hiển thị dựa vào sự phân loại đó.

Phép phân loại còn có thể được thực hiện đối với nhiều thuộc tính khác nhau của các đối tượng. Việc phân loại ghép trên cơ sở hai hay nhiều thuộc tính cho phép ta có thể nhận biết được quan hệ giữa các loại thuộc tính này trên cùng một đối tượng cũng như nhiều đối tượng lân cận nhau. Điều này có thể mang lại nhiều thông tin hơn cho người sử dụng.



**Hình 5.7:** Phân loại phường xã theo hai thuộc tính dân số và mật độ dân số

### 5.2.2. Phân lập

Hàm phân lập có mục đích chính là nhóm gộp các đối tượng, đặc điểm nào đó trong phạm vi một lớp dữ liệu. Các đối tượng vùng liền kề nhau trong lớp dữ liệu được phân lập thường có các giá trị giống nhau của một thuộc tính nào đó. Ví dụ, một lớp dữ liệu đất ngập nước có thể được tạo thành từ các đối tượng vùng của một vài lớp phụ như đất ngập nước nhiều cây gỗ, đất ngập nước cây cỏ hay vùng nước trồng. Nếu một phép phân tích đòi hỏi chúng ta phải nhận dạng chỉ những khu vực có đất của vùng đất ngập nước thì điều chúng ta mong muốn làm được đó là phân lập tất cả các đường ranh giới giữa các loại vùng đất ngập nước phân bố lân cận nhau, sau đó chỉ quan tâm đến các đường ranh giới giữa vùng có đất và vùng nước trồng.

Phép phân lập thường được sử dụng dựa vào một thuộc tính cụ thể nào đó liên quan với mỗi đối tượng, được gọi là thuộc tính phân lập. Một hay nhiều giá trị của thuộc tính phân lập sẽ được xác định và các giá trị đó thuộc vào cùng một nhóm. Trước hết, các đường đóng vai trò là đường ranh giới giữa hai đối tượng vùng được xác lập. Các giá trị của thuộc tính phân lập của các đối tượng ở bên này và bên kia đường ranh giới sau đó được so sánh với nhau. Nếu các giá trị đó là giống nhau thì đường ranh giới sẽ được xóa bỏ. Ngược lại, nếu các giá trị của thuộc tính phân lập là khác nhau ở bên này và bên kia ranh giới thì đường ranh giới sẽ được giữ nguyên.

**Trước khi phân lập**



**Sau khi phân lập**



**Hình 5.8:** Một ví dụ của phép phân lập

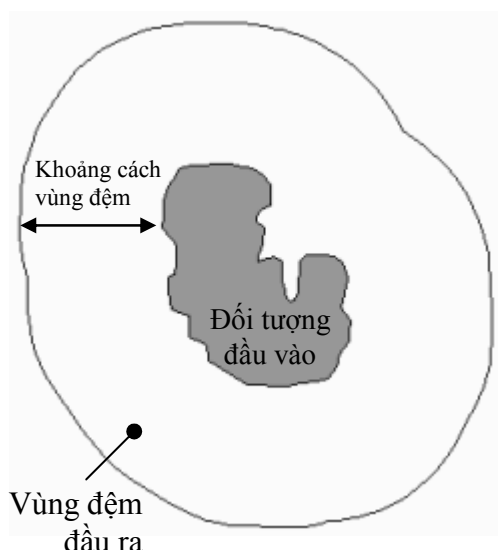
Phép phân lập rất có ích trong việc xóa bỏ bớt các thông tin không cần thiết. Giả sử sau khi ta tiến hành phân loại các đối tượng thành các lớp có kích thước

nhỏ, vừa và lớn, lúc đó, ít nhiều các đường ranh giới sẽ trở nên không cần thiết. Chính các đường ranh giới không cần thiết đó sẽ gây tốn bộ nhớ lưu trữ dữ liệu và làm chậm tốc độ xử lý của máy tính. Phép phân lập có ưu điểm là xóa bỏ được các dữ liệu địa lý hay dữ liệu bảng biểu không cần thiết, qua đó, cải thiện tốc độ xử lý của máy tính, giảm thiểu lượng dữ liệu lưu trữ cũng như sự phức tạp của hệ thống.

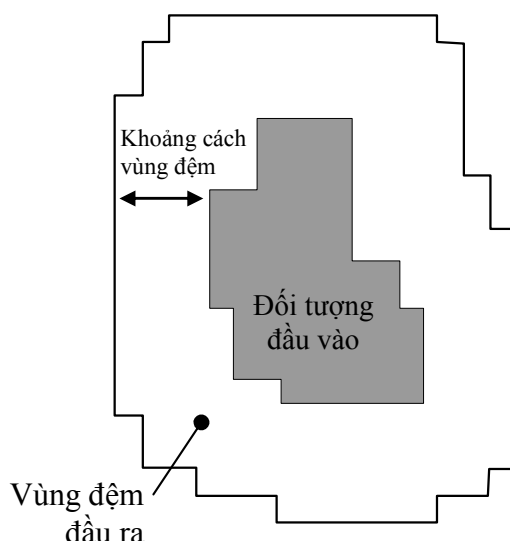
### 5.2.3. Tạo vùng đệm (buffer)

Tạo vùng đệm là một trong số các phép phân tích dữ liệu được sử dụng phổ biến. Vùng đệm là một vùng có kích thước nhỏ hơn hoặc bằng một khoảng cách được xác định trước từ một hay nhiều đối tượng. Vùng đệm có thể được xác định cho đối tượng điểm, đường hay vùng và cho dữ liệu vector hay raster. Vùng đệm thường là các vùng bên ngoài đối tượng với một khoảng cách giới hạn cho trước.

#### Vùng đệm Vector



#### Vùng đệm Raster

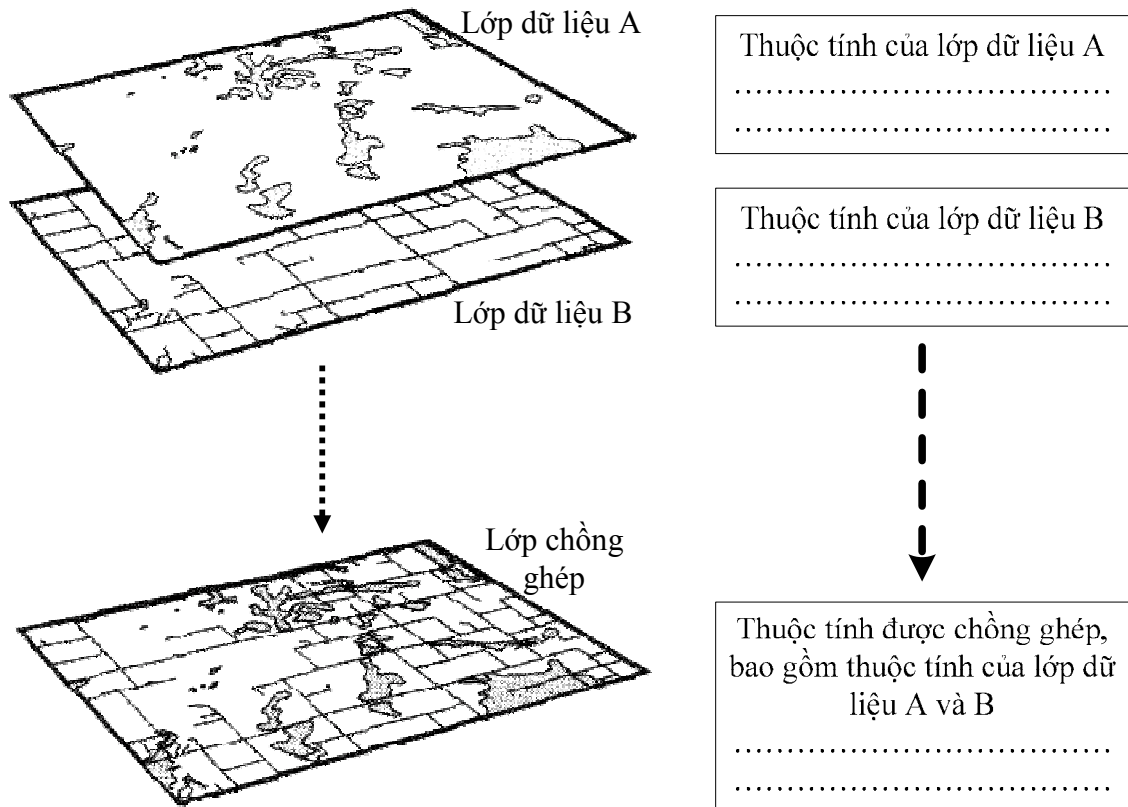


**Hình 5.9:** Ví dụ vùng đệm được tạo ra từ đối tượng vùng vector và raster

Vùng đệm thường được sử dụng phổ biến trong phân tích dữ liệu bởi lẽ rất nhiều phép phân tích không gian có liên quan đến giới hạn khoảng cách. Ví dụ như những người lập kế hoạch trong các tình trạng khẩn cấp muốn có được thông tin về khu dân cư hay ngôi trường nào nằm trong vòng bán kính 1,5 Km so với nơi có khả năng xảy ra động đất; hoặc là một người lập kế hoạch để xây dựng và phát triển hệ thống công viên cần biết được tất cả các khu đất cách xa ít nhất 10 Km so với đường quốc lộ gần nhất; hay một người chủ kinh doanh mong muốn biết được tất cả các khu vực khách hàng tiềm năng trong vòng một bán kính cho trước so với cửa tiệm của người đó... Tất cả các câu hỏi đó đều có thể được trả lời bằng cách sử dụng hợp lý công cụ tạo vùng đệm trong GIS.

### 5.2.4. Chồng ghép

Phép chồng ghép lớp bản đồ là công cụ phân tích không gian rất có lợi thế và là một yếu tố quan trọng đứng phía sau sự phát triển của công nghệ GIS. Chồng ghép chính là sự gộp chung dữ liệu không gian và thuộc tính của hai hay nhiều lớp dữ liệu và công cụ này là một trong số các phép phân tích dữ liệu phổ biến và có sức mạnh lớn trong GIS.



**Hình 5.10:** Chồng ghép dữ liệu không gian

Nhiều vấn đề trong GIS đòi hỏi sử dụng lớp chồng ghép của các dữ liệu chuyên đề khác nhau. Chẳng hạn như chúng ta muốn biết vị trí của các căn hộ giá rẻ nằm trong khu vực gần trường học; hay khu vực nào là các bãi thức ăn của cá voi trùng với khu vực có tiềm năng dầu khí lớn có thể khai thác; hoặc là vị trí các vùng đất nông nghiệp trên các khu vực đất đai bị xói mòn,... Trong ví dụ liên quan đến đất xói mòn trên, một lớp dữ liệu đất đai có thể được sử dụng để nhận biết các khu vực đất đai bị xói mòn, đồng thời lớp dữ liệu về hiện trạng sử dụng đất cũng được sử dụng để nhận biết vị trí các vùng đất sử dụng cho mục đích nông nghiệp. Thông thường thì các đường ranh giới của vùng đất bị xói mòn sẽ không trùng với các đường ranh giới của các vùng đất nông nghiệp, do đó, dữ liệu về loại đất và sử dụng đất sẽ phải được kết hợp lại với nhau theo một cách nào đó. Chồng ghép lớp bản đồ chính là phương tiện hàng đầu hỗ trợ việc thực hiện phép kết hợp dữ liệu đó.

Chồng ghép lớp có thể được xem như là một công cụ chồng lớp theo chiều thẳng đứng và hợp nhất đối với dữ liệu không gian (Hình 5.10). Các đối tượng

trong mỗi lớp dữ liệu được bố trí ở trên cùng và các đường ranh giới của các đối tượng điểm, đường và vùng được hợp nhất vào trong một lớp dữ liệu duy nhất. Dữ liệu thuộc tính cũng được ghép với nhau do vậy lớp dữ liệu mới sẽ bao gồm các thông tin chứa trong mỗi lớp dữ liệu đầu vào.

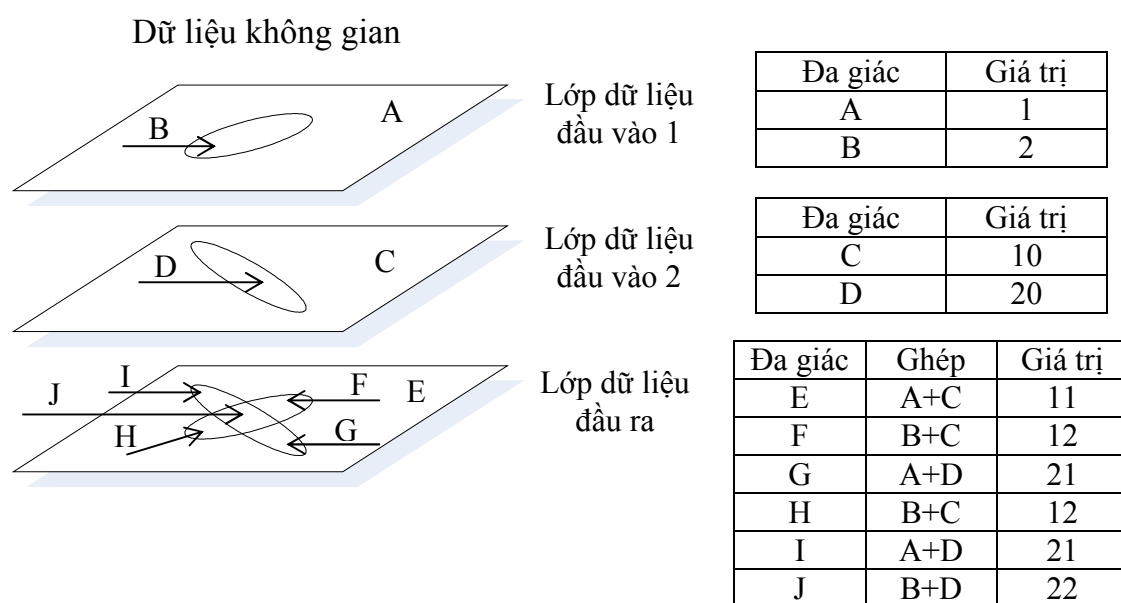
### a. Chồng ghép vector

Theo mô hình vector, các đối tượng địa lý được biểu diễn dưới dạng các điểm, đường và vùng. Vị trí của chúng được xác định bởi các cặp tọa độ và thuộc tính của chúng được ghi trong các bảng thuộc tính.

Cho đến nay, trong GIS, người ta phân biệt ba loại chồng ghép vector sau:

#### ➤ Chồng ghép đa giác trên đa giác

Chồng ghép đa giác là một thao tác không gian trong đó một lớp chuyên đề chứa các đa giác được chồng ghép lên một lớp khác để hình thành một lớp chuyên đề mới với các đa giác mới. Mỗi đa giác mới là một đối tượng mới được biểu diễn bằng một dòng trong bảng thuộc tính. Mỗi đối tượng có một thuộc tính mới được biểu diễn bằng một cột trong bảng thuộc tính.



**Hình 5.11:** Chồng ghép đa giác

Việc chồng ghép và so sánh hai bộ dữ liệu hình học có nguồn gốc và độ chính xác khác nhau thường sinh ra một số các đa giác nhỏ. Các đa giác này có thể được loại bỏ theo diện tích, hình dạng và các tiêu chuẩn khác. Tuy nhiên, trong thực tế, khó đặt ra các giới hạn để giảm được số đa giác nhỏ không mong muốn đồng thời giữ lại các đa giác khác có thể nhỏ hơn nhưng hữu ích.

Chồng ghép đa giác là một thao tác đồ sộ mà ngay cả trên các máy tính có cấu hình mạnh nhất cũng có thể đòi hỏi thời gian xử lý lâu, từ 15 đến 60 phút để chồng ghép hai tờ bản đồ trung bình. Sản phẩm của chồng ghép là một bản đồ chuyên đề bao gồm các đơn vị tương đối đồng nhất về chuyên đề và một bảng thuộc tính mở rộng.

#### ➤ **Chồng ghép điểm trên đa giác**

Các đối tượng điểm cũng có thể được chồng ghép trên các đa giác. Các điểm sẽ được gán các thuộc tính của đa giác mà trên đó chúng được chồng lên. Các bảng thuộc tính sẽ được cập nhật sau khi tất cả các điểm được kết hợp với đa giác.

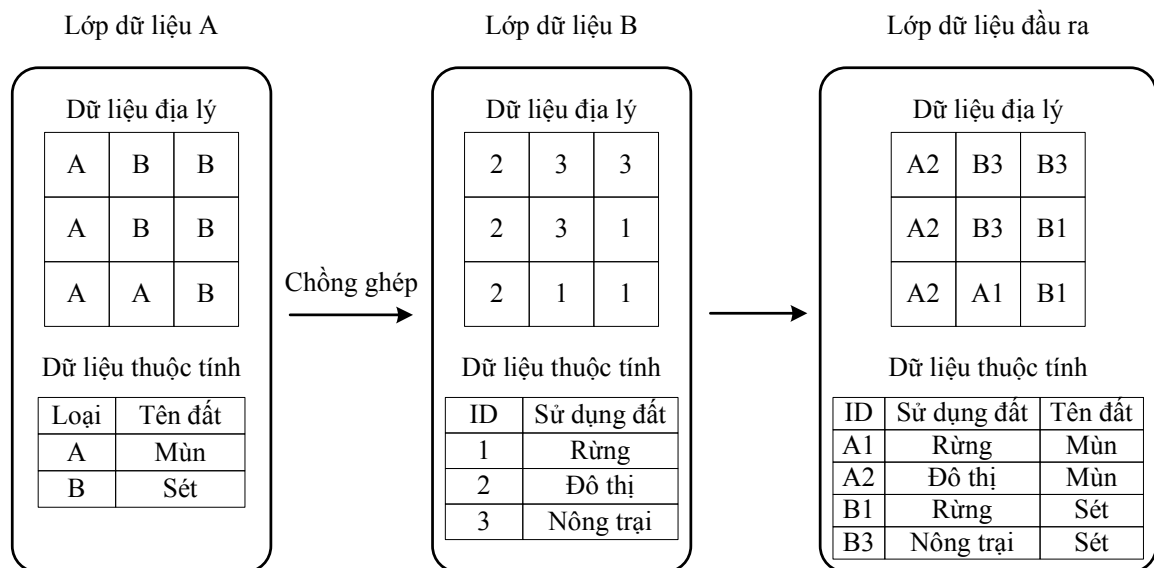
#### ➤ **Chồng ghép đường trên đa giác**

Các đối tượng đường cũng có thể được chồng ghép trên các đa giác để tạo ra một bộ các đường mới chứa các thuộc tính của các đường ban đầu và của các đa giác. Cũng như trong chồng ghép đa giác, các điểm cắt được tính toán, các nút và các liên kết được hình thành, topo được thiết lập và cuối cùng là các bảng thuộc tính được cập nhật.

### ***b. Chồng ghép raster***

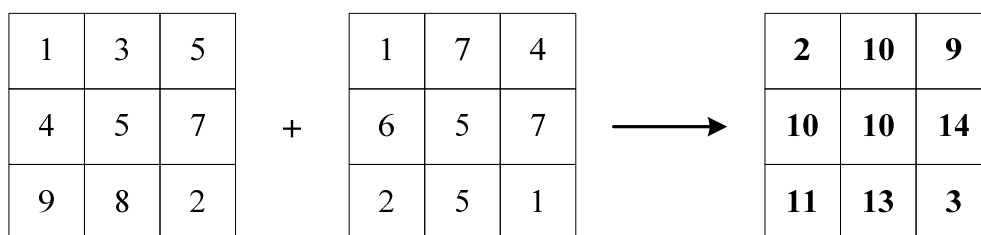
Việc chồng ghép dữ liệu hoàn toàn có thể áp dụng được đối với các dữ liệu raster và nhiều khi còn hiệu quả hơn chồng ghép vector theo như đánh giá của các nhà chuyên môn. Các vị trí của các lớp chuyên đề chỉ cần được kiểm tra xem chúng có chứa các giá trị ô lưới hay không. Trong phép so sánh ô với ô, tất cả các ô trong mỗi lớp chuyên đề đều được xem xét bất kể các giá trị của chúng. Do vậy, tổng số các ô sẽ ảnh hưởng đến thời gian xử lý của hệ thống.

Sau khi chồng ghép, các ô tổ hợp mới được hình thành với các thuộc tính bao gồm các thuộc tính từ các ô ban đầu. Và như vậy, một lớp chuyên đề mới được tạo ra. Tất cả dữ liệu raster đều bao gồm các ô lưới nên không cần phân biệt điểm, đường hay vùng. Khác với chồng ghép vector, chồng ghép raster không tạo ra các đa giác nhỏ không mong muốn, bởi lẽ dữ liệu raster bao gồm các ô lưới kích thước bằng nhau.



**Hình 5.12:** Sự kết hợp ô-ô trong phép chồng ghép raster

Trong chồng ghép raster, các thao tác số học (+, -, \*, /) và một số thao tác logic (AND, OR,...) và thống kê có thể thực hiện được một cách trực tiếp trong quá trình chồng ghép các lớp dữ liệu raster. Ví dụ hai lớp chuyên đề A và B sau có thể được cộng, trừ, nhân, chia,... để tạo ra một lớp chuyên đề C mới thông qua các thao tác toán học hay logic.



**Hình 5.13:** Minh họa thao tác chồng ghép toán học hai lớp dữ liệu raster

Để phát huy ưu điểm của chồng ghép raster, trong nhiều hệ GIS, dữ liệu vector được chuyển đổi sang dữ liệu dạng raster trước khi chồng ghép và sau đó, các kết quả chồng ghép lại được chuyển đổi sang dạng vector để tạo ra các sản phẩm đồ họa đẹp hay sử dụng tiếp trong các thao tác cần đến dữ liệu vector.

### 5.2.5. Phân tích mạng

Mạng là tập hợp các đối tượng dạng tuyến kết nối với nhau. Những đường giao thông, đường cấp điện, cấp nước, thoát nước,... là những ví dụ về mạng; mạng được dùng để di chuyển nguồn từ vị trí này đến vị trí khác.

Cơ sở của phép phân tích mạng là:

- Các mạng có tính kết nối, liên tục.



- Các quy tắc di chuyển trong mạng.
- Các đơn vị đo lường.
- Các tích tụ giá trị thuộc tính do di chuyển.
- Các quy tắc vận dụng các giá trị thuộc tính.

Phân tích mạng có rất nhiều ứng dụng trong thực tế và có thể được minh họa bằng các ví dụ khác nhau. Chẳng hạn như sự vận chuyển của nước và trầm tích trong một hệ thống sông có thể được dự báo thông qua mô hình mạng. Hay khi nhiều trận mưa lớn xuất hiện ở một khu vực, các tác động sinh ra do dòng chảy tăng lên có thể rất phức tạp; bằng cách dự báo chính xác lưu lượng nước chảy qua mạng sông suối, quy mô và vị trí của lũ lụt có thể được dự báo trước để chuẩn bị cho các hoạt động ứng cứu. Hoặc là các ứng dụng tối ưu hóa tuyến đường bao gồm từ việc chọn tuyến đường cho xe cấp cứu, cứu hỏa, xe cảnh sát cho tới đặt lịch trình hàng không, chọn tuyến dịch vụ xe khách, đưa thư, thu gom rác thải đô thị,...

Phân tích mạng phụ thuộc vào sự tồn tại của các mối quan hệ giữa các đối tượng trong hệ dữ liệu. Chẳng hạn, hệ thống phải biết được các tuyến giao thông tiếp nhận lượng phương tiện tham gia giao thông ở các chỗ giao cắt với các tuyến đường nào. Mỗi mắt xích và mỗi nút phải có một số hiệu duy nhất và chương trình phải chứa các thuyết minh về vị trí mà một con đường bắt đầu và kết thúc.

Phân tích mạng còn có thể bao gồm việc mô phỏng sự di chuyển các nguồn như ô tô, người, rác thải,... dọc theo các tuyến đường. Các nguồn này di chuyển dọc theo các con đường sẽ bắt gặp các cản trở như hạn chế tốc độ, hạn chế trọng tải, đèn giao thông, rào chắn, đường một chiều,... Do vậy, các yếu tố mạng cũng bao gồm các mắt xích, các rào cản, điểm dừng và các trung tâm. Việc gán thuộc tính cho các yếu tố đó cho phép mô phỏng các tình huống thực. Ví dụ, các thuộc tính biểu diễn sự cản trở trong mạng có thể được biểu thị bằng các đơn vị thời gian phải dừng lại ở vị trí đèn giao thông, bến phà, bến xe,... như 3 phút, 5 phút hay 1 tiếng, 5 ngày,... Mỗi khi tất cả các thuộc tính đã được phân bố, hệ thống có thể đánh giá sự di chuyển của các nguồn qua mạng. Một đơn vị đo như mét, giờ,... có thể được sử dụng để hệ thống đánh giá các phương án khác nhau và cuối cùng chọn ra tuyến đường tối ưu nhất, có ít cản trở nhất.

Trong thực tế, năng lực phân tích mạng của các phần mềm GIS hiện có là rất khác nhau. Chúng khác nhau ở quy mô và tính phức tạp của mô hình mạng, mức độ thực hiện và mức độ điều khiển tương tác.

### **5.3. Quy trình phân tích địa lý**

Trước khi bắt đầu bất cứ phép phân tích nào, bạn cần phải đánh giá được vấn đề và thiết lập được mục đích. Hãy nghĩ kỹ về quá trình sẽ thực thi trước khi đánh giá dữ liệu hay thực hiện bất cứ một quyết định nào đó; tìm xem những câu hỏi

nào cần thiết về dữ liệu và mô hình; tạo ra một thủ tục bao gồm các bước để quản lý quá trình tiến triển và phác thảo ra một mục đích cụ thể.

Một quy trình phân tích địa lý có thể bao gồm các bước sau:

Bước 1: Đặt vấn đề.

Bước 2: Chuẩn bị dữ liệu để phân tích không gian.

Bước 3: Thực hiện các thao tác không gian.

Bước 4: Chuẩn bị dữ liệu để phân tích thuộc tính.

Bước 5: Tiến hành phân tích thuộc tính.

Bước 6: Đánh giá kết quả.

Bước 7: Xác định lại và phân tích mới nếu cần.

Bước 8: Trình bày các kết quả cuối cùng.

#### ➤ **Đặt vấn đề**

Trong bước này cần phải xác định rõ mục đích và tiêu chuẩn phân tích. Ví dụ như mục đích phân tích là tìm ra địa điểm đổ chất thải rắn cho một thành phố, lúc đó, các tiêu chuẩn của bãi rác có thể là:

- Cách xa khu dân cư 2Km,
- Cách nhà máy nước 2Km,
- Nằm trên đất nông nghiệp,
- Nằm trong vùng đất sét,
- Diện tích trên 20 ha,
- Dễ ra vào bằng xe tải,...

Như vậy, các tiêu chuẩn lựa chọn có thể định tính hoặc định lượng hoặc cả hai và chúng có thể đề cập đến nhiều khía cạnh như kinh tế - xã hội, môi trường, thẩm mỹ,...

#### ➤ **Chuẩn bị dữ liệu để phân tích không gian**

Nếu bạn đã thiết kế và xây dựng thành công cơ sở dữ liệu địa lý, tại thời điểm này, tất cả các lớp dữ liệu đã sẵn sàng để có thể được phân tích. Có thể cần phải xử lý thêm về các lớp dữ liệu này hoặc có thể sau khi xem lại mục đích của phép phân tích, bạn khám phá ra là cần phải thêm vào một số thuộc tính cho cơ sở dữ liệu để thực hiện phép phân tích một cách hoàn chỉnh.

Công việc chuẩn bị cho phân tích không gian có thể bao gồm:

- Cắt vùng nghiên cứu khỏi bản đồ lớn có sẵn trong cơ sở dữ liệu.
- Ghép các mảnh bản đồ lại thành một bản đồ lớn thể hiện hết khu vực nghiên cứu.
- Biến đổi đơn vị diện tích như từ  $m^2$  sang ha...
- Tái phân loại để giảm số loại sử dụng đất không cần thiết,...

#### ➤ **Phân tích không gian**

Với dữ liệu đã được chuẩn bị, bạn có thể bắt đầu tiến hành các thao tác không gian để kết nối các lớp dữ liệu. Chúng ta đặc biệt quan tâm đến vấn đề tạo các vùng đệm xung quanh các đối tượng, thao tác trên các đối tượng không gian và tiến hành chồng ghép các vùng.

Mỗi thao tác sẽ tạo ra một lớp dữ liệu trung gian mới để xử lý tiếp. Loại và số lượng các thao tác không gian cần tiến hành tùy thuộc vào các tiêu chuẩn phân tích để đi đến kết quả mong muốn.

#### ➤ **Chuẩn bị dữ liệu để phân tích thuộc tính**

Cũng như dữ liệu không gian, dữ liệu thuộc tính cũng cần được chuẩn bị trước khi tiến hành phân tích. Điều đó có nghĩa là trước khi tiến hành phân tích dữ liệu bảng, chúng ta cần phải đảm bảo chắc chắn rằng bảng thuộc tính chứa đầy đủ các mục hay có sẵn các cột và dòng trống cần thiết để lưu trữ các dữ liệu mới sẽ được tạo ra khi phân tích.

#### ➤ **Phân tích thuộc tính**

Trên các dữ liệu thuộc tính có thể tiến hành các thao tác số học, logic và thống kê. Kết quả là tìm ra các đặc trưng thỏa mãn các tiêu chuẩn lựa chọn.

#### ➤ **Đánh giá kết quả**

Sau khi phân tích, ta tiến hành đánh giá kết quả phân tích về độ chính xác và nội dung. Câu hỏi đặt ra và cần trả lời là các kết quả phân tích có lý không? Có đáng tin cậy không? Các bản đồ đơn giản cùng với các bản báo cáo sẽ giúp cho bạn đánh giá được các kết quả.

Sau khi chắc chắn hoàn thành quá trình phân tích một cách chính xác và các điều kiện đều hợp lý thì ta có thể đi xem xét và đối chiếu kết quả ở ngoài thực địa. Nếu các kết quả đó không thể chấp nhận được so với thực tế thì ta có thể sử dụng bước này để xác định được những gì cần thay đổi và nâng cấp cách phân tích của mình. Sau đó, thực hiện lại quá trình phân tích.

#### ➤ **Xác định lại tiêu chuẩn phân tích và phân tích mới**

Cần xác định lại các tiêu chuẩn phân tích và tiến hành các phân tích mới trong trường hợp các kết quả phân tích không thể chấp nhận được hay còn có những hạn chế nhất định. GIS thực sự có ích trong vấn đề này bởi vì nó cho phép bạn dễ dàng thực hiện mới và bắt đầu ở bước thích hợp nào đó trong cả quá trình.

➤ **Trình bày các kết quả cuối cùng**

Các kết quả phân tích địa lý thường được trình bày dưới dạng các bản đồ và các báo cáo. Nội dung của bước trình bày kết quả phân tích sẽ được đề cập chi tiết ở chương tiếp theo về hiển thị và xuất dữ liệu.

## CHƯƠNG 6

# Hiển thị và xuất dữ liệu

### 6.1. Mở đầu

Với tư cách là một hệ thống, GIS có đầu vào và đầu ra. Đầu ra của GIS được thể hiện thông qua việc hiển thị và xuất dữ liệu. Đó chính là một nhu cầu, một nhiệm vụ và là một chức năng quan trọng của GIS. Sau khi đã được nhập và phân tích, dữ liệu GIS có thể được hiển thị, xuất ra dưới hình thức này hay hình thức khác tùy thuộc vào yêu cầu công việc và khả năng của hệ thống.

Việc hiển thị và xuất dữ liệu bằng GIS được thực hiện với mục đích chính:

- Kiểm tra và biên tập dữ liệu.
- Phân tích dữ liệu.
- Trình bày các kết quả phân tích và mô hình hóa để xem xét, lựa chọn và ra quyết định.

Với sự trợ giúp của máy tính và các thiết bị ngoại vi, dữ liệu GIS bao gồm dữ liệu không gian và dữ liệu thuộc tính thường được hiển thị và xuất ra dưới hai hình thức: bản mềm và bản cứng.

Thiết bị hiển thị và xuất dữ liệu GIS rất đa dạng và ngày càng được hiện đại hóa nhờ sự phát triển mạnh của công nghệ điện tử và công nghệ in.

#### 6.1.1. Màn hình

Màn hình là môi trường trao đổi bằng mắt chủ yếu của người sử dụng trong tất cả các công việc máy tính nói chung và GIS nói riêng. Màn hình cùng với các mạch điện tử gắn với nó tạo nên thiết bị hiển thị. Hiện tại, ta cần phân biệt hai loại màn hình đó là màn hình ống (CRT) và màn hình tinh thể lỏng (LCD).

Màn hình ống được dùng phổ biến trong các ứng dụng GIS vì nó tạo nhiều ánh sáng hơn, có độ phân giải cao hơn và dải màu hiển thị lớn hơn.

Các hình ảnh trên màn hình tái tạo được phục hồi đến 60 lần/giây và các hiển thị được dựa



**Hình 6.1:** Màn hình CRT



trên raster như hình ảnh tivi. Thông thường, các dữ liệu vector được chuyển sang dữ liệu raster để hiển thị trên màn hình.

Các màn hình hiển thị có thể được sản xuất bởi các hãng khác nhau và theo các kích thước khác nhau như 14, 15, 17, 19, 21,... inch. Màn hình có thể là đen trắng hay màu. Các màn hình màu sử dụng ba màu cơ bản là đỏ, lục và lam ở các cường độ khác nhau tạo ra một dải rộng các sắc màu.

**Hình 6.2: Màn hình LCD** Vị trí của con trỏ trên màn hình có thể được điều khiển bằng các phím mũi tên và các phím chức năng trên bàn phím hay bằng con chuột máy tính.

### 6.1.2. Máy in

Máy in được sử dụng để xuất dữ liệu dưới dạng bản cứng. Máy được gắn vào cổng giao tiếp của máy tính và cho phép in ra các dữ liệu không gian cũng như phi không gian.

Có nhiều loại máy in khác nhau do các hãng sản xuất khác nhau đưa ra trên thị trường và chúng được phân loại như sau:

- Máy in laser đen trắng.
- Máy in laser màu.
- Máy in kim.
- Máy in phun màu.
- Máy in phun đen trắng.

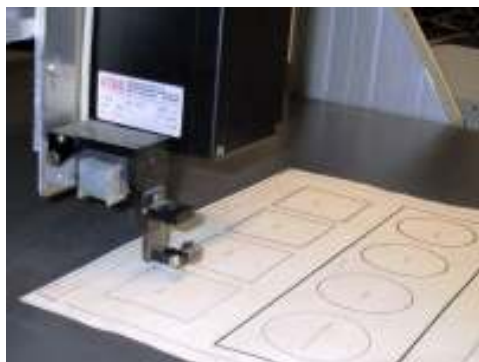


**Hình 6.3: Máy in**

### 6.1.3. Máy vẽ

Cùng với máy in, máy vẽ là một thiết bị được sử dụng phổ biến trong việc xuất dữ liệu GIS dưới dạng bản cứng, tức là biến đổi dữ liệu số sẵn dữ liệu tương tự. Các máy vẽ được cắm vào cổng nối tiếp của máy tính. Máy vẽ được chia làm các loại sau:

- Máy vẽ bút.
- Máy vẽ điện họa.
- Máy vẽ nhiệt.



**Hình 6.4:** Máy vẽ

Máy vẽ bút đôi khi còn được gọi là máy vẽ vector vì nó chỉ có thể xử lý và vẽ các dữ liệu vector. Có thể phân biệt hai loại: máy vẽ bệ phẳng và máy vẽ trục. Máy vẽ bệ phẳng có tác động cơ điện để điều khiển các chuyển động vẽ theo các hướng x và y trên môi trường vẽ được gắn cố định vào một mặt phẳng. Một hay nhiều bút được lắp vào hộp bút và có thể là bút bi hay bút máy hay bút lông có các màu sắc khác nhau. Bằng cách đó, các bản vẽ được thực hiện trên giấy, nhựa hay giấy ảnh. Máy vẽ bệ phẳng tương đối chính xác, điển hình có độ chính xác vị trí tuyệt đối là  $\pm 0,075\text{mm}$  và độ chính xác lặp lại là  $\pm 0,015\text{mm}$  ở cả hai hướng trục.

So với máy vẽ bệ phẳng thì máy vẽ trục thường được chuộng hơn vì nó có thể cho ra các bản vẽ lớn hơn và với giá thành thấp hơn. Cơ chế di chuyển bút của máy vẽ trục cũng giống như máy vẽ bệ phẳng. Môi trường vẽ được bọc một phân

xung quanh trống quay theo cả hai hướng. Độ chính xác của máy vẽ trống điển hình là +/- 0,25mm.

Ngoài ra, còn có máy vẽ điện họa bao gồm các máy vẽ điện tĩnh, điện quang và điện cảm; và máy vẽ nhiệt hoạt động theo cơ chế dùng nhiệt để tạo hình ảnh. Các máy vẽ nhiệt dùng nhiệt để sấy nóng giấy nhạy cảm với nhiệt được sơn phủ bằng hai hợp phần không màu; một khi được hâm nóng, chúng kết hợp với nhau để tạo ra một dải màu rộng.

## **6.2. Hiển thị dữ liệu**

Dữ liệu địa lý được tổ chức trong một cơ sở dữ liệu địa lý và cơ sở dữ liệu đó có thể được coi như là một bộ sưu tập dữ liệu quy chiếu không gian đóng vai trò như một mô hình về thế giới thực. Dữ liệu địa lý có hai hợp phần quan trọng là vị trí địa lý và các thuộc tính của nó.

Bảng đồ được coi là dữ liệu không gian vì thông tin nó chứa đựng liên hệ trực tiếp với các vị trí nhất định trên mặt đất. Mặt khác, dữ liệu bảng không chứa thông tin trực tiếp về vị trí mà chúng chứa thông tin mô tả. Nếu ta chỉ có bảng dữ liệu thì dữ liệu sẽ không đầy đủ vì ta không biết các đối tượng ở đâu. Nếu ta chỉ có bản đồ, ta vẫn sẽ không biết gì hơn về các đối tượng ngoài vị trí của chúng. Trong GIS, sự kết nối giữa dữ liệu không gian và thuộc tính là chìa khóa để có được thông tin đầy đủ về thế giới thực.

### **6.2.1. *Hiển thị bản đồ và bảng***

Trong GIS, việc hiển thị bản đồ và bảng hay hiển thị dữ liệu không gian và thuộc tính phụ thuộc vào từng phần mềm GIS cụ thể. Nhìn chung, có thể thực hiện công việc đó bằng các phương pháp sau:

- Dùng chuột và bàn phím.
- Dùng thực đơn (Menu) và lệnh.
- Dùng cửa sổ bản đồ.
- Dùng cửa sổ bảng.
- Dùng hộp thoại để đưa vào các tham số hiển thị.

Trong môi trường cửa sổ, việc hiển thị dữ liệu địa lý được thực hiện một cách dễ dàng nhờ vào các thực đơn, thanh công cụ và các biểu tượng. Có thể dễ dàng di chuyển, thay đổi kích thước, sắp xếp, nhìn gần, nhìn xa, cuộn ngang, cuộn dọc từng cửa sổ bảng, từng cửa sổ bản đồ để có thể biết rõ hơn về nội dung của bảng và bản đồ.



### **6.2.2. *Hiển thị bản đồ vector và raster***

Thoạt nhìn thì hai loại bản đồ này không khác nhau mấy. Sự khác nhau chỉ thể hiện rõ khi ta phóng to hay thu nhỏ. Bản đồ vector hiển thị các đường biên nhẵn giữa các đơn vị, trong khi bản đồ raster có diện mạo khối.

Một lớp dữ liệu là một bản đồ được hiển thị trong cửa sổ bản đồ, lúc đó, bạn có thể hiển thị nhiều lớp dữ liệu trong cùng một cửa sổ bản đồ. Tuy nhiên, một cửa sổ bản đồ chỉ có thể hiển thị được một bản đồ raster cùng với nhiều bản đồ vector.

Nội dung của bản đồ được xác định bởi miền của bản đồ. Cùng một miền có thể được sử dụng bởi một bản đồ vector, một bản đồ raster và một bảng. Có thể có miền lớp, miền giá trị, miền ảnh và miền nhận diện. Miền lớp là một danh sách các tên lớp. Miền giá trị là các giá trị đo đạc, tính toán hay nội suy. Miền ảnh là các giá trị phản xạ; trong ảnh vệ tinh hay ảnh quét, miền ảnh nằm trong khoảng giữa 0 và 255. Miền ảnh thực chất là một kiểu đặc biệt của miền giá trị. Miền nhận diện là mã duy nhất đối với mỗi thực thể trong bản đồ.

Bản đồ vector và raster có thể được hiển thị đồng thời trong một cửa sổ bản đồ. Dùng chức năng tùy chọn quản lý lớp dữ liệu, bạn có thể chọn các lớp cần hiển thị là các lớp mà từ đó bạn cần thu thập thông tin. Bạn có thể thay đổi các tùy chọn hiển thị lớp bản đồ cũng như có thể bổ sung hay loại bỏ các lớp bản đồ.

### **6.3. *Xuất dữ liệu dưới dạng bản đồ***

Đây chính là hình thức xuất dữ liệu dưới dạng bản cứng được dùng phổ biến để trình bày các kết quả phân tích và mô hình hóa không gian GIS. Bản thân GIS không phải là một hệ thống lập bản đồ tự động, song với GIS, ta không chỉ có thể nhập, lưu trữ và phân tích bản đồ mà còn có thể tạo ra được các bản đồ để trình bày và phục vụ quá trình ra quyết định. Như vậy, bản đồ vừa là đầu vào, vừa là đầu ra của GIS. Vấn đề thiết kế và tạo bản đồ bằng GIS phải được xem như là quá trình xây dựng đầu ra của GIS. Một bản đồ hoặc một bản báo cáo được thiết kế chuẩn giúp cho chúng ta có được một ấn tượng tốt về kết quả của dự án GIS. Ngoài ra, chúng còn làm cho người khác tăng mức độ tin cậy và dễ dàng chấp nhận kết quả của công việc cần đến sự trợ giúp của GIS.

Khả năng xuất dữ liệu dưới dạng bản đồ bằng GIS phụ thuộc vào phần cứng và phần mềm GIS. Phần mềm vector GIS thường có ưu thế hơn so với phần mềm raster GIS trong việc tạo ra các sản phẩm đồ họa đẹp nhờ sử dụng mô hình dữ liệu vector.

Trong phần tiếp theo sẽ giới thiệu phương pháp để xây dựng từng thành phần của một bản đồ, bao gồm:

- Thiết kế các thành phần của bản đồ.
- Sử dụng các ký hiệu một cách hiệu quả nhất.

- Xác định mục đích của bản đồ.
- Định nghĩa các tham số bản đồ.
- Thiết kế bản đồ kết quả.
- Chuẩn bị ký hiệu.
- Tạo bản đồ cuối cùng.

Một bản đồ thường chứa một số lớp dữ liệu kết nối với nhau để đưa ra được một sản phẩm cuối cùng. Một bản đồ bao gồm thông tin mô tả giúp cho người đọc có thể đọc được các thông tin mà bản đồ muốn thể hiện.

Các thành phần chính của bản đồ có thể được chia thành hai loại: các đối tượng địa lý và các yếu tố bản đồ.

#### ➤ *Các đối tượng địa lý*

Các đối tượng địa lý của bản đồ bao gồm đối tượng điểm, đường và vùng được tổ chức và vẽ thành các lớp khác nhau trong cơ sở dữ liệu địa lý.

Các đối tượng vùng là các vùng khác nhau trên Trái đất, chẳng hạn như vùng đất nông nghiệp, công nghiệp, vui chơi giải trí,... Đường biên của các vùng được vẽ bằng các ký hiệu đường. Các vùng có thể được tô màu dựa trên các thuộc tính của chúng và bằng các tính chất như màu, mẫu hay cả hai. Chúng cũng có thể được gán nhãn với các thuộc tính sử dụng các ký hiệu văn bản.

Các đối tượng đường là các đối tượng tuyến tính như đường giao thông, đường cấp nước, thoát nước,... Các đường được vẽ bằng các ký hiệu đường và có thể được gán nhãn với các thuộc tính sử dụng các ký hiệu văn bản.

Các đối tượng điểm là các điểm hoặc các điểm nhãn của vùng. Chúng được vẽ với các ký hiệu đánh dấu hoặc được gán nhãn với các thuộc tính sử dụng ký hiệu văn bản.

#### ➤ *Các yếu tố bản đồ*

Các yếu tố bản đồ giúp cho người đọc có thể dễ dàng xem các thông tin trên bản đồ hơn. Đó là các yếu tố:

- Các đầu đề và các đoạn văn bản miêu tả mục đích của bản đồ, chúng được thể hiện bằng các ký hiệu văn bản.
- Các đường biên của bản đồ được vẽ bằng các ký hiệu đường.
- Các chú giải mô tả các ký hiệu được sử dụng để thể hiện các yếu tố địa lý thông qua các ký hiệu về đường, mẫu, các ký hiệu đánh dấu và văn bản.

- Yếu tố để chỉ hướng Bắc và các thanh tỷ lệ mô tả hướng và tỷ lệ của bản đồ. Chúng được vẽ bằng các ký hiệu đường, mẫu và các ký hiệu văn bản.

### ❖ Sử dụng các ký hiệu

Các đối tượng địa lý cũng như các yếu tố bản đồ đều được vẽ bằng các ký hiệu khác nhau. Phần này sẽ mô tả sơ lược về các đặc tính của ký hiệu được sử dụng để tạo bản đồ.

#### ➤ Các tham số xác định ký hiệu

Các ký hiệu được xác định thông qua một loạt các tham số. Các tham số chung nhất đối với hầu hết các ký hiệu bao gồm: màu sắc, mẫu và kích thước.

##### ✓ Màu sắc

Màu sắc là một tham số định tính. Các màu sắc sẵn có phụ thuộc vào thiết bị mà bạn sử dụng để hiển thị các ký hiệu. Một máy vẽ bút có thể chỉ cho phép 4 hoặc 8 bút vẽ (số màu phụ thuộc vào số bút vẽ). Trái lại, một màn hình đồ họa có thể hiển thị 16 màu hoặc nhiều hơn tại một thời điểm; trong khi các máy vẽ điện tử có thể cho phép hiển thị đến hàng trăm màu trên bản đồ.

##### ✓ Mẫu

Các mẫu để chỉ sự lặp lại của các phân tử trong ký hiệu. Ví dụ một đường chấm liên tục có thể được hợp thành bởi các điểm được tách rời nhau theo từng đoạn nhất định.

.....

hoặc cũng có thể là một chuỗi hai điểm một được tách rời nhau

.. .. .

##### ✓ Kích thước

Kích thước quyết định độ rộng và chiều cao của các ký hiệu văn bản hay dấu hiệu hoặc quyết định độ rộng của ký hiệu đường

nhỏ      vừa      lớn      \*      \*      \*

\_\_\_\_\_      \_\_\_\_\_      \_\_\_\_\_

Trên đây chỉ là một số tham số chung nhất để xác định ký hiệu bản đồ. Ngoài ra, còn có các tham số khác để xác định một số ký hiệu, chẳng hạn như đối với các ký hiệu văn bản, ngoài các tham số màu sắc, mẫu, kích thước, còn có thêm các tham số như:

Khoảng trống: “a b c d e f” hay “abcdef”

Phông chữ: **abcdef** **abcdef** *abcdef* ...

Và kiểu: gạch chân hay *in nghiêng* hay **in đậm** ...

➤ Các file ký hiệu

Trong phần mềm GIS thường có sẵn các bộ ký hiệu. Người sử dụng có thể lựa chọn các bộ ký hiệu đó để lập bản đồ hoặc tự tạo các ký hiệu riêng của mình bằng việc số hóa các ký hiệu đã vẽ hay in trên giấy hay dùng phần mềm để xây dựng các ký hiệu riêng.

➤ Kiểu ký hiệu

Có bốn kiểu ký hiệu:

- Kiểu tô màu,
- Kiểu đường,
- Kiểu đánh dấu,
- Kiểu văn bản.

❖ **Sử dụng các ký hiệu một cách hiệu quả nhất**

Việc sử dụng các ký hiệu có hiệu quả hay không sẽ ảnh hưởng đến việc kết nối thông tin với bản đồ. Dưới đây là các hướng dẫn để hiển thị các bản đồ một cách có hiệu quả nhất.

➤ Sử dụng màu

Màu sắc là một yếu tố rất quan trọng trong quá trình xây dựng bản đồ. Do vậy, phải rất lưu ý trong khi chọn màu và sắc màu. Nên chọn các màu thích hợp cho các đối tượng, ví dụ như màu đỏ để thể hiện vùng cấm, màu xanh để thể hiện vùng được phép sử dụng và màu vàng là vùng chờ xem xét,... Không nên chọn quá nhiều màu để tránh làm méo mó thông tin và làm cho bản đồ thêm rắc rối.

➤ Sử dụng các mẫu

Phải chọn các mẫu sao cho tăng được khả năng phân biệt các vùng với nhau trên cả bản đồ và chú giải. Một số mẫu rất khó phân biệt so với các mẫu khác và nên tránh sử dụng chúng.

Phải luôn cẩn thận khi chọn các mẫu đường hoặc các mẫu chấm chấm,... bởi vì các mẫu này có thể rất khó nhận ra. Ngoài ra, cũng phải xem xét thiết bị xuất dữ liệu là thiết bị gì để lựa chọn cho phù hợp. Ví dụ như nếu bản đồ sẽ được vẽ ra trên một máy vẽ bút thì các đường liền nét sẽ dễ thể hiện hơn là các đường chấm chấm,

bởi vì khi vẽ các đường liền nét, sự di chuyển của bút vẽ chỉ nằm trên đường đó trong khi vẽ các đường không liền nét thì bút vẽ phải liên tục được nhấc lên và đặt xuống. Sử dụng bút vẽ nhiều cho các đường không liền nét có thể làm giảm độ bền và gây hư hỏng bút do mực không thể chảy xuống liên tục được. Ngoài ra, cũng cần nên tránh sử dụng những mẫu tô hoặc các đường quá dày đặc. Làm như vậy có thể tốn rất nhiều thời gian khi vẽ và có thể làm hư hại đến bút vẽ và thiết bị vẽ.

➤ Sử dụng các ký hiệu văn bản

Nói chung, các ký tự văn bản thường được sử dụng trong đầu đề, chú thích và ngay trong bản đồ như các nhãn về địa danh, tên các đối tượng. Ngoài ra, nó cũng có thể được sử dụng để ghi nguồn dữ liệu, chương trình tạo và ngày tạo bản đồ.

Kích cỡ của các ký tự và kiểu ký tự văn bản có thể làm cho thông tin nổi bật hơn. Đối với các đầu đề của bản đồ thường dùng ký tự văn bản có kích thước lớn, còn nguồn dữ liệu hoặc các phần khác thường sử dụng ký tự nhỏ hơn. Hãy chú ý rằng các ký tự to, đậm sẽ mất thời gian để vẽ hơn so với các ký tự nhỏ và không đậm bằng.

### ❖ Các bước để tạo bản đồ bằng GIS

Các bước trong quá trình tạo bản đồ bằng GIS do ESRI đưa ra bao gồm:

- Bước 1: Xác định mục đích và yêu cầu của bản đồ.
- Bước 2: Xác định kích thước và tỉ lệ của bản đồ.
- Bước 3: Thiết kế khung bản đồ.
- Bước 4 Chuẩn bị dữ liệu.
- Bước 5: Tạo bản đồ cuối cùng.

➤ **Bước 1:** Xác định mục đích và yêu cầu của bản đồ

Thông thường, trước khi tạo bản đồ, bạn cần phải trả lời được ba câu hỏi sau: “Tại sao phải xây dựng bản đồ?”, “Bản đồ dành cho ai?” và “Bản đồ sẽ được thể hiện như thế nào?”. Mục đích chính là xác định được phạm vi của bản đồ, thiết lập một danh sách các đối tượng và từ đó phát triển được một phương án cụ thể để xây dựng bản đồ.

*Khả năng truyền đạt của bản đồ:* Bản đồ là một phương tiện để truyền đạt thông tin. Nó có thể rất đơn giản nhưng vẫn chứa đựng đầy đủ thông tin tổng quát. Các bản đồ thường được thiết kế cho một mục đích nhất định. Ví dụ như bản đồ giao thông nhấn mạnh đến các đối tượng giao thông, bản đồ dân số chú trọng đến vấn đề dân số, còn bản đồ đất đai thì mô tả vị trí của các kiểu đất khác nhau.

*Khả năng của người đọc bản đồ:* Khả năng nhận thức của chúng ta phụ thuộc rất nhiều vào kiến thức và kinh nghiệm của bản thân. Mỗi người có kỹ năng đọc bản đồ khác nhau. Sự thành thạo trong việc đọc và hiểu bản đồ thay đổi theo các nhóm người khác nhau. Nội dung của bản đồ thường phụ thuộc rất nhiều vào người sử dụng và mục đích sử dụng. Phải luôn xem xét đến khả năng của người đọc; một số người có thể đọc bản đồ không tốt lắm do vậy các lớp bản đồ phải rõ ràng và chỉ nên bao gồm những thông tin cần thiết nhất mà thôi.

➤ **Bước 2:** Xác định kích thước và tỉ lệ bản đồ

Các thông số kích thước và tỉ lệ bản đồ phụ thuộc vào mục đích và độ già của bản đồ.

*Thiết bị và kích thước bản đồ:* Việc chọn thiết bị để cho ra bản đồ phụ thuộc vào mục đích và ý tưởng của bản đồ; nó cũng phụ thuộc vào những thiết bị nào có thể có được. Quyết định lựa chọn thiết bị và kích thước sẽ ảnh hưởng rất lớn đến chất lượng của bản đồ kết quả. Hiện thị kết quả trên màn hình đẹp chưa chắc đã in ra giấy đẹp.

Thiết bị xuất ra có thể được phân loại theo chức năng và chất lượng. Màn hình thường được sử dụng để kiểm tra kết quả. Sử dụng màn hình, ta có thể xem được rất nhanh những vùng có diện tích nhỏ, các bản đồ đơn giản và để hiển thị các kết quả đồ họa mà thông qua đó bạn có thể hỏi đáp về một số thông tin. Màn hình có thể xem là tương tự như máy vẽ phun mực.

Sau khi thiết kế xong bản đồ cuối cùng, có thể tạo một file vẽ riêng. Các máy vẽ bút và máy vẽ điện tử có thể sử dụng file này. Máy vẽ bút là một thiết bị sử dụng các bút màu để in ra được những văn bản và yếu tố đồ họa có chất lượng khá cao; còn máy vẽ điện tử thì có thể in ra được những hình ảnh có chất lượng rất cao. Ngoài ra, còn có một số tùy chọn đối với bản đồ ra, chẳng hạn như tạo các file có định dạng khác nhau.

Khi nghiên cứu để tìm ra một kích thước tối ưu và tỉ lệ dài rộng thích hợp của bản đồ, người ta nhận thấy rằng hầu hết các bản đồ có tỉ lệ dài rộng là 4:3 thì tốt nhất. Kích thước của bản đồ thường phụ thuộc vào mục đích của bản đồ. Một bản đồ để trưng bày thường có kích thước lớn, trong khi đó bản đồ nằm trong tài liệu báo cáo có thể bị giới hạn bởi kích thước của trang giấy trong tài liệu đó.

*Tỉ lệ:* Vì những lý do thực tế, các bản đồ thường là mô tả bề mặt Trái đất theo một tỉ lệ nhất định nào đó. Bạn không thể thể hiện tất cả các thông tin lên trên bản đồ theo ánh xạ 1:1 được, do vậy cần phải có một tỉ lệ đủ để lưu trữ những dữ liệu chủ yếu và thể hiện nó như là một bản đồ theo một tỉ lệ nhất định.

Chọn tỉ lệ thích hợp cho bản đồ phụ thuộc vào độ phân giải của dữ liệu cũng như độ chi tiết của bản đồ muốn thể hiện.

Một bản đồ có thể chứa nhiều biến diễn địa lý ở các tỉ lệ khác nhau, ví dụ như trên bản đồ có thể có thêm một hay nhiều bản đồ để chỉ một vị trí xác định nào

đó. Điều quan trọng là phải chỉ ra các tỉ lệ của bản đồ tương ứng với vị trí của chúng.

Cần chú ý rằng khi phóng to bản đồ dựa trên kích thước cơ sở của nó có thể gây ra kết quả là làm thiếu hụt các chi tiết cũng như độ phân giải trên bản đồ. Ngược lại, giảm tỉ lệ của bản đồ có thể dẫn đến sự nhầm lẫn vì thông tin trên bản đồ bị nhỏ lại và rất khó phân biệt.

➤ **Bước 3:** Thiết kế khung bản đồ

Tính cân xứng: Phải sắp xếp các thành phần của bản đồ như thế nào đó để tạo nên được tính cân xứng. Trong bất cứ thiết kế đồ họa nào cũng nên luôn luôn nghĩ đến người đọc sẽ tập trung vào đâu. Sử dụng màu sắc, mẫu, ký hiệu hợp lý sẽ gây được sự chú ý và làm cho người đọc dễ theo dõi thông tin hơn.

Cách dễ nhất để thể hiện được thông tin và tạo ra sự tương xứng trước khi xây dựng bản đồ là phác thảo ra được thiết kế sơ bộ về khung của bản đồ. Hãy thay đổi vị trí đầu đề bản đồ, thanh tỉ lệ và chú giải đến khi nào bạn cảm thấy đã có được một sự tương xứng giữa các yếu tố đó.

Không có một phương pháp tổng quát nào để sắp xếp các phần tử của bản đồ. Thông thường, bằng cách thử nghiệm các phương án, bạn mới có thể tạo ra được sự tương xứng nhất trong bản đồ. Có thể bạn sẽ thấy mất thời gian khi thiết kế các bản đồ và tạo ra được tính tương xứng nhưng nhờ vào đó bạn có thể tạo ra được một bản đồ dễ đọc, dễ hiểu, dễ nhớ và gây được ảnh hưởng lớn về kết quả phân tích của mình.

➤ **Bước 4:** Chuẩn bị dữ liệu cho bản đồ

Trước khi xây dựng bản đồ, bạn cần phải chuẩn bị các file dữ liệu sẽ được sử dụng trong bản đồ. Hầu hết việc chuẩn bị liên quan đến vấn đề chọn các ký hiệu cho các đối tượng khác nhau trên bản đồ. Trong bước này có hai nhiệm vụ chính sau đây:

- Xác định các ký hiệu cần sử dụng để vẽ đối với từng đối tượng trong một lớp.
- Tạo các file chú giải chứa các ký hiệu và các văn bản xuất hiện trong bản đồ.

*Vẽ các đối tượng sử dụng thuộc tính của chúng*

Có hai cách để xác định các ký hiệu. Cách thứ nhất là xác định các ký hiệu trước khi vẽ đối tượng. Ví dụ như bạn có thể xác định trước là sẽ dùng màu đỏ để vẽ các con đường; lúc đó, tất cả các con đường trong lớp đường sẽ được vẽ bằng màu đỏ. Điều này có thể chấp nhận được nếu như bản đồ đơn thuần chỉ để hiển thị vị trí của các con đường. Trong trường hợp bạn có một thuộc tính phân biệt các con đường đã được nâng cấp (đã được lát nhựa) và các con đường chỉ mới được

nâng cấp một phần (chỉ mới được rải sỏi) và bạn muốn thể hiện các con đường theo thuộc tính đó. Lúc này, bạn phải sử dụng thuộc tính để vẽ đối tượng bằng cách dùng ngay cột lưu trữ thuộc tính cần sử dụng trong bảng thuộc tính hoặc có thể thêm một mục vào trong bảng thuộc tính để lưu trữ giá trị ký hiệu cho từng đối tượng.

Cách thứ hai là sử dụng các thuộc tính chỉ tới một bảng tìm kiếm hay bảng tra để xác định các ký hiệu. Bảng tìm kiếm là một file dữ liệu trong đó chứa một mục giống như mục trong bảng thuộc tính đối tượng và một mục khác gọi là SYMBOL (biểu tượng). Đối với mỗi giá trị thuộc tính duy nhất, sẽ có một bảng ghi tương ứng trong bảng tìm kiếm với số hiệu nhất định của ký hiệu. Sử dụng bảng tìm kiếm có một số thuận tiện. Trước hết, bạn có thể sử dụng bất cứ ký hiệu nào để thể hiện thuộc tính không bị hạn chế bởi giá trị được lưu trong bảng thuộc tính. Ngoài ra, bạn cũng có thể thay đổi các ký hiệu một cách dễ dàng hơn; và bạn cũng có thể dễ dàng thay đổi mục trỏ đến bảng tìm kiếm bằng cách thay đổi tên mục thuộc tính trong bảng tìm kiếm. Sử dụng bảng tìm kiếm sẽ tiết kiệm được không gian lưu trữ; thay vì phải lưu trữ mỗi số hiệu của ký hiệu cho từng bản ghi trong bảng thuộc tính, nay bạn chỉ cần lưu trữ ký hiệu cho mỗi giá trị của mục trong bảng tìm kiếm.

#### *Chú giải*

Khi bạn sử dụng các ký hiệu để vẽ các đối tượng trên bản đồ, bạn nên thêm các chú giải cho các ký hiệu được sử dụng trên bản đồ. Để thực hiện điều này, bạn cần phải tạo ra một file chứa các chú giải xác định các ký hiệu đó. File chú giải là một file văn bản bình thường có thể được tạo ra bằng chương trình soạn thảo văn bản nào đó. Trong file này, bạn phải xác định các ký hiệu trong chú giải và đoạn văn bản mô tả tương ứng với ký hiệu đó.

#### ➤ **Bước 5:** Tạo bản đồ cuối cùng

Sau khi tất cả các công việc chuẩn bị đã được hoàn thành, bạn có thể bắt đầu tạo bản đồ trên máy tính bằng các lệnh thích hợp và bản đồ tạo ra có thể được đưa ra máy in hoặc máy vẽ.

## **6.4. Chuẩn dữ liệu và chất lượng dữ liệu**

Trong GIS, chất lượng dữ liệu là một đề tài rất đáng quan tâm bởi lẽ dữ liệu là một hợp phần, vừa là đầu vào, vừa là đầu ra của GIS. Chất lượng đầu vào sẽ ảnh hưởng đến chất lượng của đầu ra. Lời cảnh báo “rác vào rác ra” đối với máy tính hoàn toàn có thể áp dụng cho GIS. Điều đó nói lên tầm quan trọng của dữ liệu mà chúng ta thu thập và đưa vào trong máy tính để tạo nên cơ sở dữ liệu địa lý. Nếu ta đưa vào các dữ liệu chất lượng thấp, có nhiều sai sót thì thật khó có thể cho ra được các dữ liệu, thông tin mới chất lượng cao.

GIS là một công cụ trợ giúp ra quyết định mạnh mẽ nhưng đắt. Nó đắt không chỉ vì giá thành phần cứng, phần mềm và nhân lực được đào tạo mà còn vì giá thành



của các quá trình thu thập, nhập, quản lý và phân tích dữ liệu. Quá trình lựa chọn và ra quyết định qua GIS chỉ có thể được cải thiện khi dữ liệu được thu thập, nhập, lưu trữ và phân tích là đủ tin cậy và không có sai sót đối với các mục đích đặt ra.

### **6.4.1. Vấn đề chuẩn dữ liệu**

Chất lượng dữ liệu địa lý thường được xem xét, đánh giá sau khi các quyết định không đúng được đưa ra và thường mất mát tài chính hay những đụng chạm đến con người đã xuất hiện. Nhà sản xuất thông tin địa lý phải chịu trách nhiệm khi sản phẩm của họ bị phát hiện có sai sót, được thiết kế dở hay được dùng theo cách và vì các mục đích không được các nhà thiết kế chủ định.

Các chuẩn dữ liệu được xác định, kiểm tra và báo cáo đúng có thể bảo vệ cả người sản xuất lẫn người sử dụng thông tin địa lý. Khi dữ liệu được cung cấp theo khuôn dạng chuẩn và ở mức chất lượng được xác định và chấp thuận, người sản xuất được bảo vệ khỏi trách nhiệm trong trường hợp sử dụng không đúng. Các tiêu chuẩn đó cũng bảo vệ người dùng khỏi dựa vào thông tin không đúng.

GIS cung cấp phương tiện để thông tin, dữ liệu địa lý được dùng cho hàng loạt ứng dụng và bởi các người dùng với kỹ năng khác nhau. Do đó các dữ liệu cần dùng trong các quá trình ra quyết định phải có chất lượng dự báo được hay biết được.

Các chuẩn dữ liệu phải đáp ứng nhu cầu của người dùng. Như vậy, cộng đồng người dùng phải được can dự trực tiếp trong việc xác định rõ các chuẩn dữ liệu đối với cơ sở dữ liệu GIS và trong việc giải quyết các ràng buộc thực tế như ngân sách, năng lực kỹ thuật hay giá thành sản phẩm. Các chuẩn chất lượng và các phương pháp đo lường chất lượng phải được xác định rõ ràng trước khi bắt đầu nhập dữ liệu. Chất lượng dữ liệu cần được đo lường định kỳ khi phát triển và cập nhật cơ sở dữ liệu của hệ thống.

### **6.4.2. Các yếu tố chất lượng dữ liệu**

Để mô tả hay đánh giá chất lượng dữ liệu, người ta sử dụng một loạt các yếu tố chất lượng như độ chính xác về vị trí, độ chính xác về thuộc tính, độ chính xác về thời gian, sự nhất quán logic, sự đầy đủ và độ phân giải, lịch sử và việc sử dụng dữ liệu. Các chỉ tiêu chất lượng được sử dụng để mô tả các yếu tố chất lượng.

#### **➤ Độ chính xác về vị trí**

Độ chính xác về vị trí là độ lệch của vị trí địa lý của đối tượng trên bản đồ so với vị trí thực của nó trên mặt đất. Trong đo vẽ bản đồ thông thường, độ chính xác tỉ lệ thuận với tỉ lệ bản đồ. Bản đồ tỉ lệ 1:1000 sẽ chính xác hơn bản đồ tỉ lệ 1:100.000.

Trong số hóa thủ công, độ chính xác của dữ liệu số phụ thuộc vào mật độ các điểm dùng để biểu diễn đối tượng. Khi số hóa các đường hay đa giác, độ chính xác tăng khi khoảng cách giữa các điểm được số hóa giảm.

Ngoài ra, độ chính xác của dữ liệu số phụ thuộc vào số con số có nghĩa mà máy tính cho phép lưu trữ. Các chuyển đổi dữ liệu từ vector sang raster và ngược lại cũng dẫn đến sự giảm độ chính xác về vị trí.

Để đảm bảo độ chính xác về vị trí, có thể sử dụng các phương pháp khác nhau như đo đạc độc lập, đánh giá chủ quan, điều chỉnh,... trong đó, đo đạc độc lập là phương pháp thường được sử dụng nhất.

### ➤ **Độ chính xác về thuộc tính**

Trong thực tế, độ chính xác của các dữ liệu thuộc tính cũng quan trọng như độ chính xác của các dữ liệu vị trí. Dữ liệu thuộc tính sai sẽ dẫn đến những sai sót trong dữ liệu cuối cùng; ví dụ như khi các định nghĩa không đầy đủ về các loại đối tượng dẫn đến các đối tượng bị phân loại nhầm lẫn.

Các sai số thuộc tính có thể là sai số thô, ngẫu nhiên hay có hệ thống và có thể phát sinh do thiết bị và người vận hành nó.

Nhìn chung, độ chính xác của các thuộc tính số được kiểm tra bằng cách so sánh dữ liệu với các dữ liệu thực được gán ngẫu nhiên và được biểu thị bằng độ lệch chuẩn và sai số hệ thống.

Độ chính xác của dữ liệu phân loại được thể hiện bằng phần trăm các phân loại đúng hoặc sai như: 99% các đối tượng được phân loại đúng.

Biện pháp đảm bảo chất lượng ở đây bao gồm các đo đạc độc lập, đo lại, điều chỉnh, đánh giá chủ quan,...

### ➤ **Độ chính xác về thời gian**

Thời gian đóng vai trò quan trọng trong các dữ liệu địa lý. Tính thời sự hay độ chính xác về thời gian của dữ liệu đề cập đến hai vấn đề sau:

- Dữ liệu hình học và thuộc tính của các đối tượng sẵn có được thay đổi khi nào?
- Các đối tượng mới với các hình thể và thuộc tính mới xuất hiện khi nào?

Mức độ thời sự cần thiết phụ thuộc vào loại đối tượng và ứng dụng. Ví dụ, các ứng dụng tiện ích như điện, nước cần các dữ liệu cập nhật hai đến ba tuần một lần về ranh giới nhà cửa, đường sá,... trong khi các ranh giới phân vùng thì thay đổi chậm hơn và như vậy ít phải cập nhật thường xuyên.

Tính thời sự của dữ liệu trở nên đặc biệt quan trọng khi nhiệm vụ đặt ra cần đến dữ liệu từ nhiều nguồn khác nhau. Tính thời sự thường được thể hiện ở ngày chụp không ảnh, ngày quét ảnh vệ tinh hay ngày đo đạc,...

### ➤ **Sự nhất quán logic**

Các quan hệ logic trong dữ liệu phải thỏa mãn những yêu cầu đặt ra về các mối quan hệ giữa các đối tượng dựa trên các tác vụ cần thực hiện. Các đa giác liền kề nhau phải có các ranh giới chung, các đường phải nối với nhau theo mạng logic, các đa giác phải được khép kín.

Sự nhất quán của các quan hệ logic thường khó đo lường. Trong bất kỳ trường hợp nào, chất lượng phụ thuộc vào việc sử dụng các chương trình kiểm tra khi dữ liệu được đưa vào.

Các quan hệ logic trong dữ liệu bản đồ là thiết yếu đối với người sử dụng GIS. Người sử dụng hệ biên tập bản đồ có thể làm việc tốt với các dữ liệu spaghetti, tuy nhiên, topology kém chất lượng có thể sẽ làm phức tạp hoặc gây sai sót trong nhiều chức năng GIS như chồng ghép hay phân tích mạng.

### ➤ **Sự đầy đủ và độ phân giải**

Sự đầy đủ dữ liệu nói lên sự có mặt của tất cả các dữ liệu không gian và thuộc tính cần thiết. Có thể kiểm tra sự đầy đủ bằng cách so sánh số các đối tượng trong dữ liệu được đưa vào và số các đối tượng thực tế trong khu vực liên quan. Kết quả so sánh có thể được biểu thị bằng phần trăm các đối tượng thiếu vắng. Nhìn chung, càng nhiều thuộc tính được đưa vào thì sự mô tả càng tốt.

Độ phân giải chỉ thị mật độ quan trắc như kích thước ô mạng trong mô hình raster GIS, kích thước của đa giác trong vector GIS, hay sự ưu tiên của các điểm đo đạc. Nói cách khác, độ phân giải là đơn vị nhỏ nhất có thể phân biệt được hay đơn vị nhỏ nhất được biểu diễn trên bản đồ.

Sự đầy đủ và độ phân giải cũng cần được khai báo đối với dữ liệu được đưa vào là lưu trữ để thông báo cho người sử dụng về các hạn chế liên quan.

### ➤ **Lịch sử dữ liệu**

Thông tin về lịch sử hay nguồn gốc, gốc gác, lai lịch của dữ liệu bao gồm tên tổ chức sản xuất dữ liệu, mục đích sản xuất, ngày sản xuất, nơi sản xuất,...

### ➤ **Việc sử dụng dữ liệu**

Đây cũng được xem như là một yếu tố chất lượng dữ liệu. Thông tin về việc sử dụng dữ liệu bao gồm cơ quan sử dụng, mục đích sử dụng, các hạn chế sử dụng,...

### **6.4.3. Các nguồn sai sót**

Sai sót là khó tránh khỏi trong quá trình sử dụng GIS và nó làm giảm chất lượng của dữ liệu GIS. Do đó, cần hiểu rõ các nguồn sai sót có thể để tìm cách hạn chế chúng đến mức tối đa.

➤ **Nguồn sai sót độc lập với xử lý GIS**

a. *Thiết bị*

- Hệ thống viễn thám vệ tinh
- Máy ảnh hàng không
- Máy thu GPS
- Các thiết bị đo các giá trị thuộc tính khác nhau

b. *Sai sót trong quá trình lập bản đồ*

- Lập bản đồ
  - Sai số trong tính toán và mạng lưới trắc địa
  - Sự không chính xác của công cụ lập bản đồ
  - Sự không chính xác trong vẽ bản đồ
- Biên tập dữ liệu
  - Tính toán
  - Phóng to, thu nhỏ và vẽ lại

c. *Sai sót do những thay đổi ở thực địa*

- Các đối tượng đã đăng ký thay đổi thuộc tính
- Các hiện tượng mới phát sinh

d. *Sai sót do thiếu độ phủ hay độ phân giải*

➤ **Sai sót trong xử lý dữ liệu GIS**

a. *Nhập dữ liệu*

- Lỗi do số hóa
  - Lỗi thiết bị
  - Lỗi người thao tác
- Sự không chính xác trong khi vào dữ liệu thuộc tính
  - Lỗi do con người (thiếu các chương trình kiểm tra)

b. *Trữ dữ liệu*

- Độ chính xác số của máy tính không đủ
- Lỗi của môi trường trữ

*c. Lỗi trong thao tác dữ liệu*

- Raster sang vector
- Vector sang raster
- Tổng quát hóa và làm mỏng
- Kết hợp các lớp
- Chồng ghép
- Nội suy
- Phân tích các dữ liệu viễn thám và các dữ liệu khác

*d. Lỗi trong trình bày dữ liệu*

- Lỗi thiết bị
- Lỗi của môi trường trình bày

➤ **Sai sót về phương pháp**

- Sai sót về các phương pháp dùng để thu thập dữ liệu
- Mật độ quan trắc không đủ
- Các lớp và các đối tượng xác định không tốt
- Sự thiếu kinh nghiệm của người biên tập dữ liệu
- Ranh giới không chắc chắn giữa các vùng

#### **6.4.4. Kiểm tra và sử dụng dữ liệu**

Có thể kiểm tra độ chính xác của dữ liệu qua sự đối chiếu với các dữ liệu khác chính xác hơn. Các dữ liệu đối chiếu đó có thể là các số liệu đo đạc ngoài thực địa hay các bản đồ tỉ lệ lớn hơn. Việc kiểm tra có thể tiến hành ở trong phòng hay ở ngoài thực địa.

Dữ liệu có thể bị sử dụng sai do không hiểu rõ hay do thiếu kinh nghiệm. Để dữ liệu đúng dữ liệu cần có các mô tả rõ ràng về độ chính xác của dữ liệu cũng như kinh nghiệm chuyên môn của người sử dụng trong ứng dụng liên quan. Một số quy tắc sau có thể giúp kiểm tra và sử dụng đúng dữ liệu:

- Dùng các chương trình kiểm tra để bảo đảm chất lượng.
- Kiểm tra dữ liệu càng sớm càng tốt.
- Kiểm tra dữ liệu ở nhiều giai đoạn thao tác dữ liệu.
- Không xáo trộn các dữ liệu có độ chính xác cao và dữ liệu có độ chính xác thấp.
- Hiểu biết bản chất của dữ liệu, dù nó là dữ liệu không gian hay phi không gian.
- Có đầu óc phê phán trong mọi sử dụng dữ liệu.
- Dùng các kết quả xử lý dữ liệu một cách thận trọng.
- Nêu sự bất chính xác gắn với các kết quả và các phân tích.

## CHƯƠNG 7

# Các phát triển mới trong GIS

### 7.1. GPS - Bản đồ di động

Công nghệ GPS sẽ ngày càng được phát triển mạnh mẽ với sự tích hợp nhiều công nghệ mới để gia tăng số lượng, chất lượng dữ liệu và hiệu quả đo đạc; đồng thời chi phí của các thiết bị này có xu hướng giảm để có thể tiếp cận được nhiều người dùng.

Công nghệ GPS có thể được tích hợp vào các thiết bị di động như điện thoại và các thiết bị PDAs (Personal Digital Assistants). Công nghệ này cung cấp các dịch vụ định vị cho người dùng điện thoại di động, cho phép định hướng và phân tích địa lý thông qua các thiết bị kết nối không dây.

Tác động của công nghệ không dây sẽ đáng kể nhất khi các thông tin định vị chính xác được cần đến một cách nhanh chóng hay thường xuyên và liên tục. Khi được kết hợp với công nghệ GPS, các máy tính xách tay và các phần mềm, dữ liệu không gian và các công nghệ không dây khác có khả năng hỗ trợ rất tốt cho người sử dụng trong các trường hợp nhất định như các dịch vụ khẩn cấp, cấp cứu, các hoạt động đảm bảo an toàn quân sự hay cộng đồng, hay thậm chí trong trường hợp điều khiển ô tô và người điều khiển cần sự hỗ trợ về hướng đi và đường đi đúng.

Ngoài ra, GPS cũng được tích hợp với các thiết bị quét địa hình bằng laser để mang lại hiệu quả trong việc thu thập các thông tin chi tiết liên quan đến dữ liệu không gian. Trong ứng dụng này, các thiết bị quét ba chiều được phát triển để có thể đo đạc được vị trí theo trục thẳng đứng và theo phương ngang của đối tượng.

### 7.2. Cải thiện công nghệ viễn thám

Việc thu thập dữ liệu không gian về căn bản sẽ được cải thiện cùng với nhiều tiến bộ liên tục trong việc thu thập và xử lý ảnh viễn thám. Số lượng vệ tinh nhiều, độ phân giải về không gian và thời gian được cải thiện và nền sensor (đầu cảm biến) mới sẽ góp phần làm tăng lượng dữ liệu có thể thu thập được. Chúng ta sẽ có

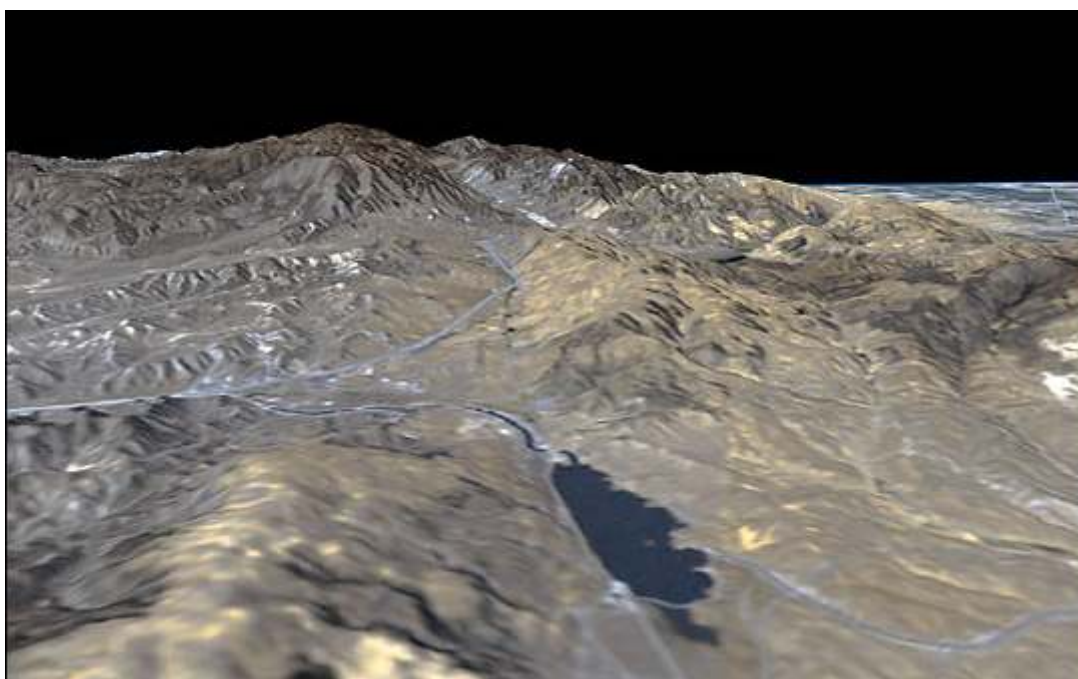


*Hình 7.1: Minh họa thiết bị di động được tích hợp công nghệ GPS*

thể phán đoán được các hiện tượng sẽ xảy ra và cũng có thể xác định vị trí của các đối tượng đã được đo đạc từ trước nhưng với độ chính xác được cải thiện đáng kể.

Thiết bị bay con thoi thực hiện nhiệm vụ vẽ bản đồ địa hình bằng sóng rada (SRTM - Shuttle Radar Topography Mission) là một ví dụ của việc cải thiện công nghệ viễn thám. Một thiết bị quét bằng rada được cài đặt trên thân của máy bay con thoi này vào tháng 2 năm 2000. Thiết bị quét đã thu thập được khoảng 85% dữ liệu độ cao địa hình của bề mặt Trái đất. Các dữ liệu này sẽ được xử lý trong vài năm đến và sẽ đem lại các dữ liệu địa hình đồng bộ trên phạm vi toàn cầu và có thể được ghép với các dữ liệu không gian khác.

Chất lượng dữ liệu được cải thiện trên phạm vi toàn thế giới là một trong số các lợi thế chính của dữ liệu SRTM. Dữ liệu SRTM tại hầu hết các vùng của Châu Á, Nam Mỹ và Châu Phi thực tế là tốt hơn nhiều so với các dữ liệu khác.



**Hình 7.2:** Dữ liệu mới của SRTM giúp cải thiện việc phân tích không gian

Sự tiến bộ trong lĩnh vực máy bay thu nhỏ và kích thước thật được biết đến như các phương tiện bay có người lái từ xa (RPVs - Remotely Piloted Vehicles) hay các phương tiện không gian không người lái (UAVs - Unmanned Aerial Vehicles) có thể dẫn đầu trong việc gia tăng tính khả thi của các sản phẩm không ảnh phạm vi rộng lớn hơn. NASA cũng như các phòng thí nghiệm của chính phủ khác và các hãng không gian vũ trụ đã và đang phát triển các phương tiện bay không người lái để thu thập ảnh và dữ liệu liên quan đến khí quyển và bề mặt Trái đất.





**Hình 7.3:** Máy bay không người lái tầm xa Altus (NASA)



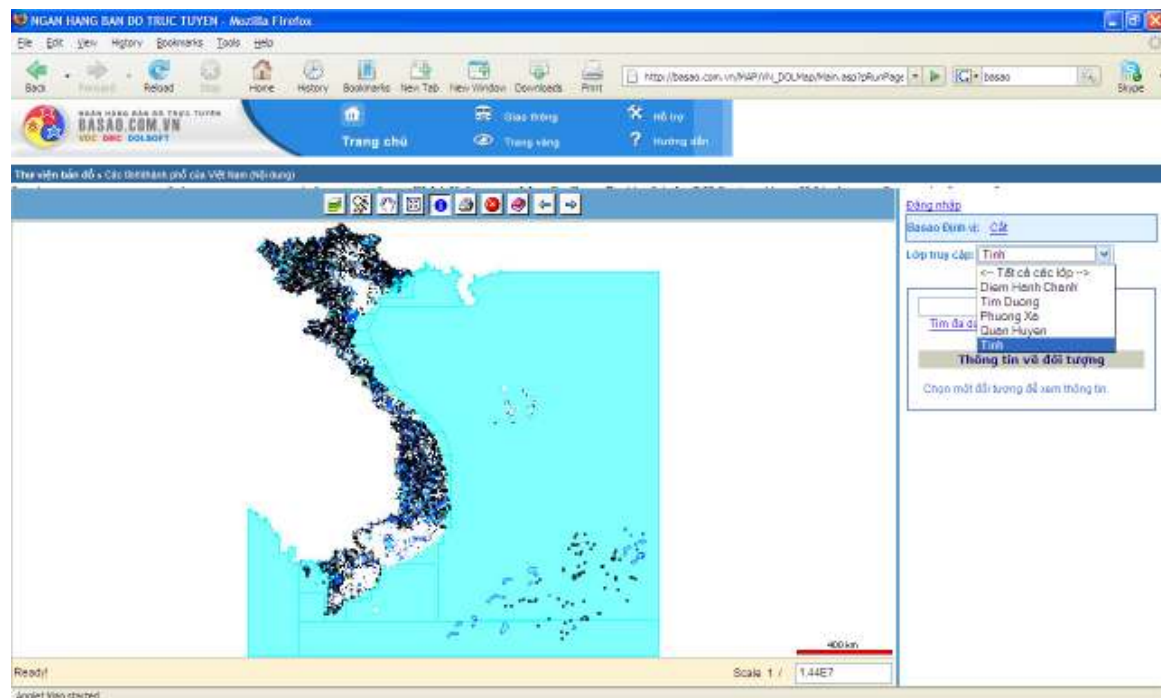
**Hình 7.4:** Thiết bị bay không người lái được trang bị máy ảnh để thu thập không ảnh ở độ cao trung bình và thấp

### 7.3. Bản đồ Internet

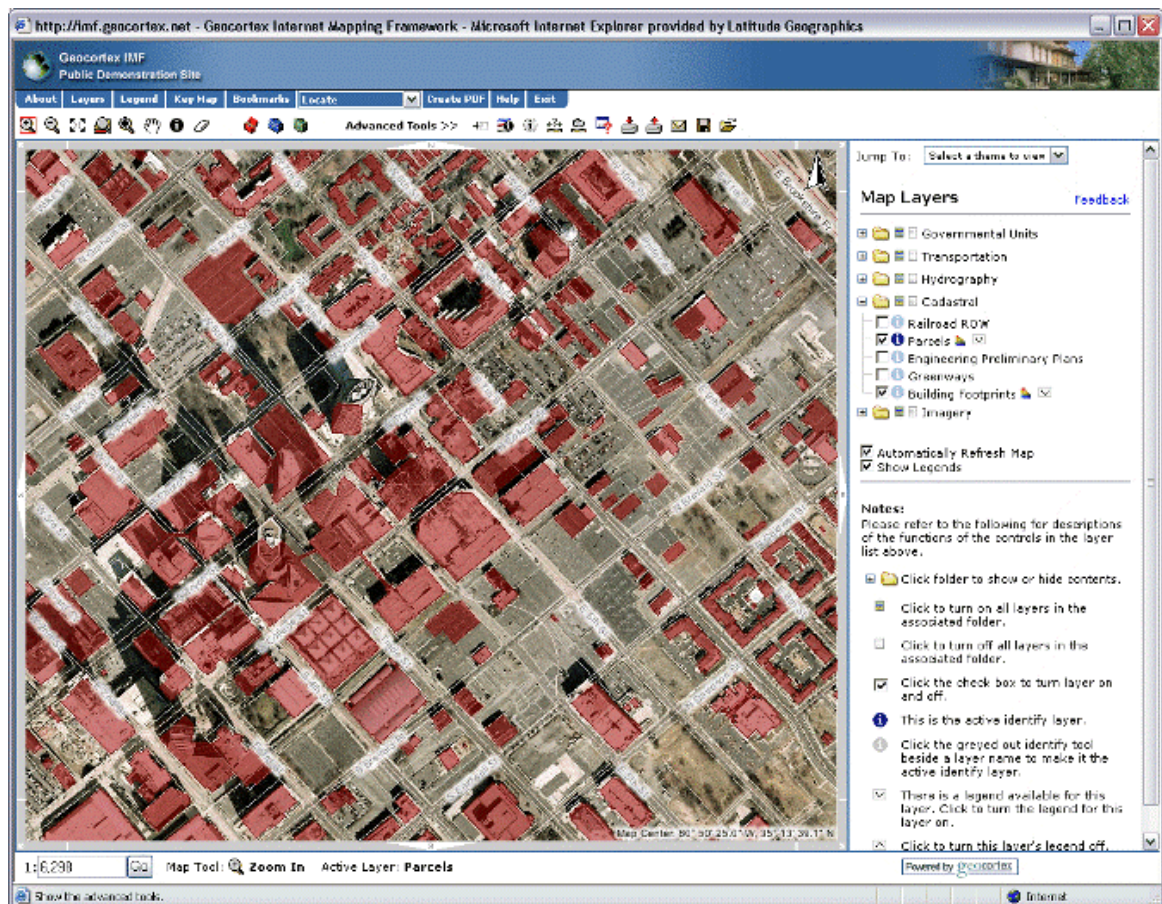
Internet sẽ có một tác động quan trọng ngày càng lớn đối với GIS; đặc biệt là trong việc mở rộng số lượng và bề rộng của dữ liệu không gian và cải thiện việc cập nhật và phân phối thông tin không gian. Các bản đồ đang ngày càng phổ biến trên các trang web trên thế giới. Bản đồ internet cung cấp các chức năng bản đồ thông thường chẳng hạn như các hỗ trợ về phương hướng để biểu hiện trạng thái không gian hay cung cấp hình ảnh ảo của các vị trí không gian.

Các ứng dụng bản đồ internet cho phép người sử dụng tạo các bản đồ trên các trang web. Những người sử dụng riêng lẻ được trang bị các nút điều khiển các lớp dữ liệu được hiển thị, phạm vi của vùng trên bản đồ và các biểu tượng được sử dụng để biểu hiện đối tượng bản đồ. Các ứng dụng bản đồ internet phải được thiết kế đặc biệt phù hợp khi một số lượng lớn người sử dụng cần truy cập một số lớp dữ liệu giới hạn để biên soạn bản đồ. Người sử dụng internet có thể lựa chọn chủ đề, các biến số và biểu tượng khác với bản đồ tĩnh mặc định nhưng trong phạm vi các định nghĩa về đặc tính của bản đồ mà người vẽ bản đồ của trang web tạo ra.

Đa số các ứng dụng bản đồ internet được xây dựng cho đối tượng sử dụng với ít hiểu biết về dữ liệu, bản đồ cũng như phân tích không gian. Chính việc đơn giản hóa này đã hạn chế số lượng các phép toán, công cụ không gian có thể cho phép sử dụng.







*Hình 7.5: Các ứng dụng cung cấp dịch vụ bản đồ dựa vào internet*

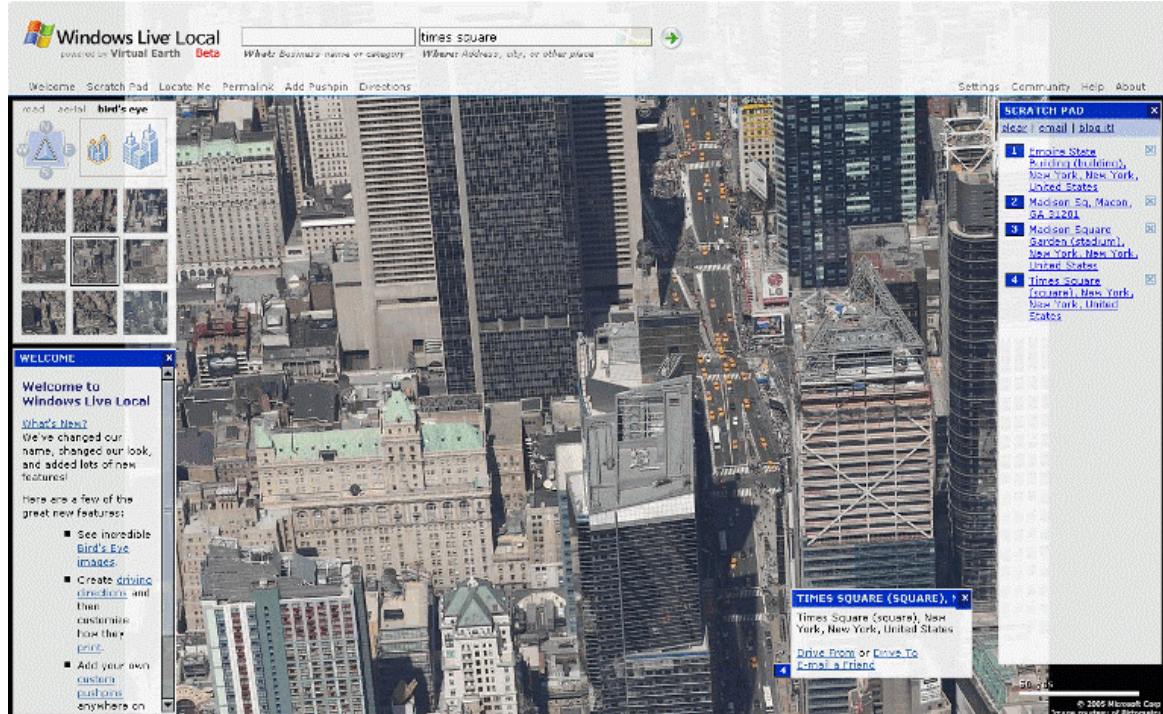
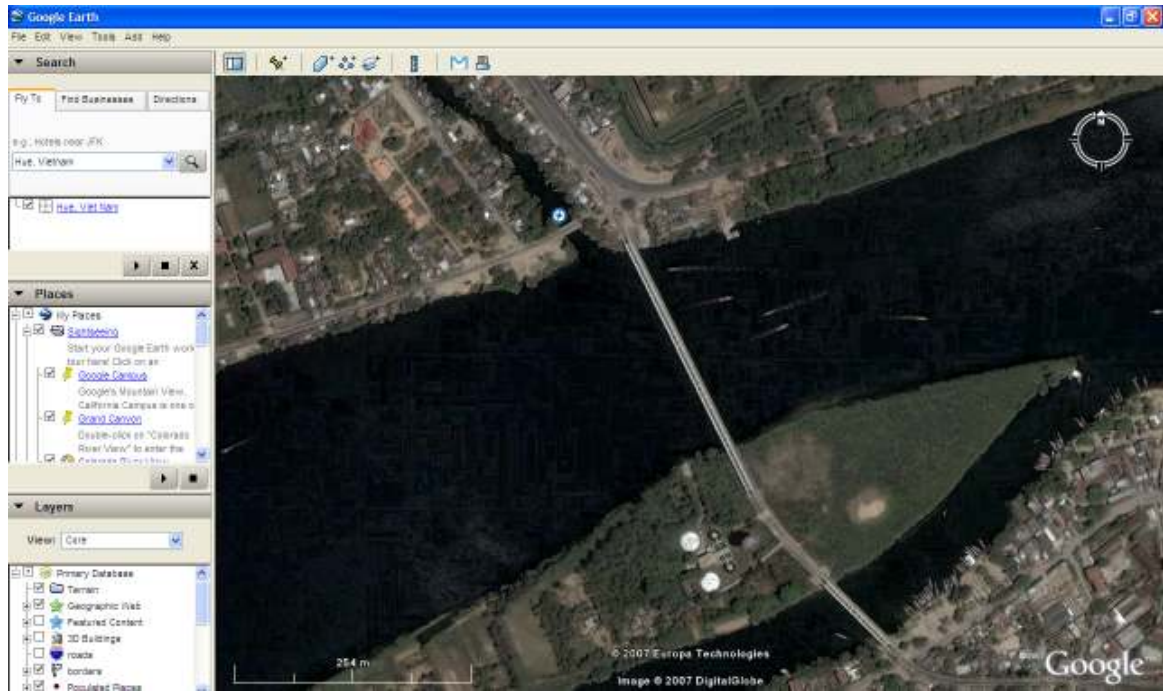
## 7.4. Thế giới ảo

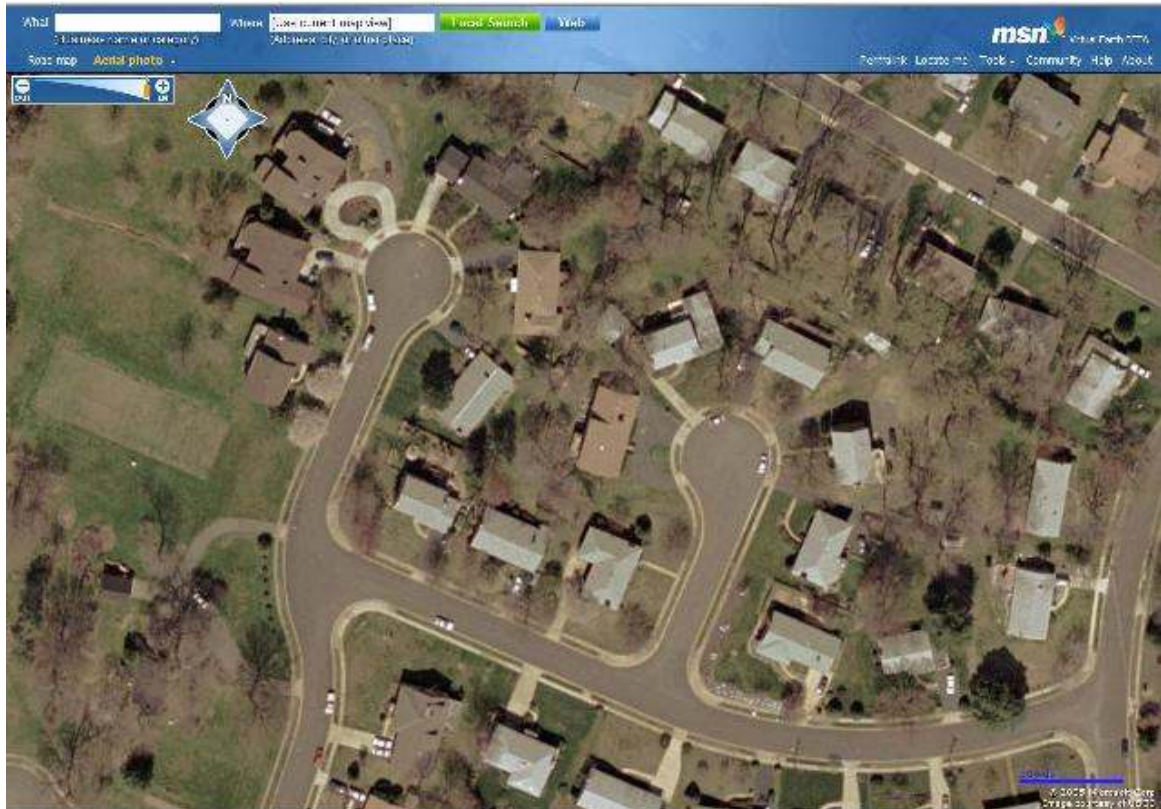
Dữ liệu không gian đang được sử dụng ngày càng nhiều trong các ứng dụng về thế giới ảo. Thế giới ảo thực chất là sự biểu diễn các thực thể có thật bằng máy tính. Các thực thể này có thể là nhà cửa, con người, các phương tiện đi lại hay các thực thể khác. Bất cứ một đối tượng thực hay tưởng tượng nào đều có thể được mô tả trong thế giới ảo.

Thông thường thì thế giới ảo biểu diễn hai chiều nhưng biểu diễn đó tạo cảm giác ba chiều cho người xem. Độ bóng khác nhau trên bề mặt các đối tượng, bóng của chúng và nghệ thuật phối cảnh, tạo chiều sâu, tất cả đều được sử dụng để tạo ra cảm giác chiều sâu cho người xem.

Dữ liệu không gian trong thế giới ảo đang được sử dụng rộng rãi trong các hoạt động có giá thành đắt, tốn nhiều thời gian hay không khả thi trong việc đưa người đến các vị trí địa hình thực tế. Đào tạo phi công là một ví dụ điển hình; thế giới ảo được sử dụng rộng rãi trong đào tạo phi công bởi vì một chuyến bay ảo bằng mô hình sẽ có chi phí thấp hơn nhiều để xây dựng và hoạt động mô hình so với một chuyến bay thực tế. Do vậy, có thể tiết kiệm được chi phí rất lớn trong việc đào tạo này.







*Hình 7.6: Dữ liệu không gian được sử dụng trong các ứng dụng về thế giới ảo*

## 7.5. GIS mở

### ➤ Các chuẩn mở cho GIS

Các chuẩn mở thường có tác dụng làm giảm bớt các trở ngại trong việc chia sẻ dữ liệu và thông tin. Dữ liệu không gian có cấu trúc rất phức tạp, có thể phức tạp hơn so với các loại dữ liệu khác. Dữ liệu có thể là raster hay vector, kiểu thực hay kiểu nhị phân và được mô tả bằng các đối tượng điểm, đường hay vùng. Ngoài ra, các phần mềm khác nhau có thể lựa chọn cách thức để ghi các ảnh raster bằng các định dạng khác nhau và dữ liệu có thể được phân phát sử dụng các phương tiện vật lý khác nhau hay được định dạng theo các cách khác nhau. Nếu một người sử dụng nhập một file ảnh theo một định dạng nào đó nhưng máy tính của người đó không hỗ trợ các phương tiện để làm việc với định dạng dữ liệu được nhập hoặc thậm chí không hiểu cấu trúc file, lúc đó người sử dụng sẽ không thể sử dụng dữ liệu như mong muốn. Việc các hệ thống không tương thích với nhau được mô tả như là không có khả năng trao đổi dữ liệu, và các chuẩn mở nhằm đến mục đích xóa bỏ rào cản trao đổi dữ liệu đó.

Các chuẩn mở có mục đích cung cấp một khung dữ liệu chung trong việc mô tả, xử lý và chia sẻ dữ liệu. Các chuẩn mở cũng nhằm đến việc cung cấp các phương pháp cho người cung cấp và người sử dụng để xác nhận sự chính xác so với tiêu chuẩn.

Các chuẩn mở trong GIS vẫn còn mới lạ. Tuy đã có rất nhiều các nhà cung cấp phần mềm, các nhà phát triển dữ liệu và các tổ chức chính phủ và giáo dục là thành viên của hội GIS mở; nhưng một số thành phần của các bộ chuẩn vẫn đang được xây dựng và hoàn thiện. Tầm quan trọng của việc tuân theo các chuẩn GIS mở sẽ được thấy rõ ràng hơn trong tương lai không xa.

### ➤ *Nguồn GIS mở*

Nguồn phần mềm mở khác với hầu hết các phần mềm khác là chúng được cung cấp miễn phí với mã nguồn. Có rất nhiều đặc điểm khác nhau về nguồn phần mềm mở trong các định nghĩa chính thức được đưa ra bởi các tổ chức cung cấp dịch vụ; tuy nhiên, các đặc điểm chính xung quanh khái niệm nguồn phần mềm mở đó là mạng lưới mở của các cộng tác viên, các tài liệu và mã nguồn chia sẻ miễn phí.

Có nhiều nguồn phần mềm mở khác nhau về chủng loại cũng như cấu hình hệ thống phù hợp, và bao gồm cả GIS. Các nguồn này chủ yếu được cung cấp thông qua các hệ thống mạng internet hay mạng nội bộ. Có nhiều tổ chức sử dụng nguồn GIS mở vì các sản phẩm thương mại có thể không hỗ trợ các chức năng hay khả năng cần thiết. Tuy nhiên, các phát triển của nguồn GIS mở có thể tạo nên nhiều thuận lợi cho việc có được các nguồn tài nguyên hiệu quả nhưng chi phí thấp hơn nhiều so với việc đầu tư mua các sản phẩm thương mại.