

MỘT SỐ VẤN ĐỀ VỀ PHƯƠNG PHÁP LUẬN METADATA CỦA CÁC HỆ THỐNG THÔNG TIN ĐỊA LÝ

Phạm Thị Bích Liên

Trung Tâm Công Nghệ Thông Tin Địa Lý, Trường Đại học Bách Khoa

TÓM TẮT

Trong hệ thống thông tin địa lý, nhiều người cùng làm ra dữ liệu và cùng sử dụng chung tài nguyên dữ liệu có sẵn trên toàn hệ thống. Sự hữu hiệu của hệ thống được khẳng định khi có nhiều người cùng dùng chung dữ liệu. Mỗi người dùng khi sử dụng dữ liệu của hệ thống thường đòi hỏi những thông tin về dữ liệu. Do đó, mỗi nguồn dữ liệu khi cung cấp cho hệ thống đều được kèm theo những thông tin liên quan đến dữ liệu. Những thông tin về dữ liệu được gọi là metadata. Ở đây, một số yêu cầu về chuẩn nội dung metadata được thảo luận.

ABSTRACT

In a geographic information system, a lot of people make data together and share the available data resources on the whole system. The system is effective when there are many people sharing data. Users of the geographic information systems request the information of the data when they take it for some target. Delivered on the network, data must be attached the information concerning it, called metadata. Here, some issues of metadata content standard is discussed.

1. GIỚI THIỆU

Dữ liệu là thành phần quan trọng nhất và tốn nhiều chi phí nhất (80% chi phí) trong một hệ thống thông tin địa lý. Hiểu một cách chi tiết về nguồn dữ liệu là rất cần thiết nếu như dữ liệu được dùng để làm cơ sở cho việc ra quyết định. Những thay đổi trong dữ liệu GIS có thể gây khó khăn trong việc tìm kiếm và một người sử dụng dữ liệu sẽ không thể dễ dàng xác định những thay đổi gì đã xảy ra chỉ bằng cách hiển thị những thành phần không gian và bảng thuộc tính của dữ liệu.

Hơn nữa, cùng một loại dữ liệu không gian nhưng có thể được sử dụng cho nhiều mục đích khác nhau, việc xây dựng lặp lại dữ liệu sẽ gây tốn kém và không hiệu quả. Do đó cần thiết phải có sự trao đổi, chia sẻ để tận dụng các nguồn tài nguyên dữ liệu có sẵn.

Metadata có thể giúp tổ chức tạo ra dữ liệu bảo quản giá trị của dữ liệu và giữ cho chúng tiếp tục được sử dụng qua nhiều năm, giúp cho người sử dụng tìm kiếm được nguồn dữ liệu thích hợp để sử dụng.

Song song với việc tạo ra dữ liệu địa lý của các hệ thống thông tin địa lý (GIS), những nhà kiến tạo hệ thống cần phải xây dựng đầy đủ các nội dung của metadata. Ở đây, những vấn đề về phương pháp luận cũng như nội dung metadata của một hệ thống thông tin địa lý cần quan tâm trong tiến trình xây dựng cơ sở dữ liệu được trình bày.

2. ĐỊNH NGHĨA METADATA

Theo nghiên cứu của Ủy ban dữ liệu địa lý liên bang Hoa Kỳ, từ Metadata có nguồn gốc từ tiếng Hy Lạp như từ Metamorphosis (sự biến hình, sự biến hóa). Meta có nghĩa là sự thay đổi và metadata hay còn gọi là 'data about data' mô tả nguồn gốc và dấu vết của sự thay đổi dữ liệu không gian địa lý.

Có nhiều định nghĩa và mục đích sử dụng cho Metadata. Trong GIS, Metadata được sử dụng như những tư liệu cung cấp nội dung và điều kiện của dữ liệu.

Theo *The International Organization for Standardization Technical Committee 211, Geographic information and Geomatics (1994)*: "Metadata là dữ liệu về nội dung, chất lượng, điều kiện và những đặc điểm khác của dữ liệu"

Theo *Environmental Systems Research Institute (ESRI) (1995)*: "Metadata được xem như là những thông tin diện có hệ thống và suy diễn về nội dung, cấu trúc, mối quan hệ của dữ liệu cũng như sự mô tả và phạm vi sử dụng của dữ liệu được lưu trữ trong cơ sở dữ liệu".

Theo *The United Nations Environment Program/ Global Resource Information Database (UNEP/GRID) (1996)*: "Metadata là dữ liệu tham khảo về nội dung và vị trí của dữ liệu và cách lưu trữ thông tin. Metadata chứa những thông tin tổng quát về một tập dữ liệu riêng biệt hoặc nơi cung cấp phương pháp truy cập đến dữ liệu đó"

Theo *The Consortium for International Earth Science Information Network (CIESIN) (1996)*: "Metadata là thông tin mô tả tập dữ liệu, bao gồm hướng dẫn sử dụng dữ liệu, mô tả về thư mục lưu trữ, kiểu mẫu, bảng tóm tắt về dữ liệu và bất kỳ thông tin nào thêm vào đều được yêu cầu định rõ mối quan hệ với những thông tin có sẵn"

Theo *Committee for Earth Observing Satellites (CEOS) (1996)*: "Metadata được xem như là dữ liệu về dữ liệu được chứa đựng trong các tập dữ liệu và cung cấp cách hiểu về nội dung và tính hữu dụng của tập dữ liệu"

Các nội dung của một Metadata gồm có:

- Thông tin nhận dạng dữ liệu là những thông tin cơ bản về tập dữ liệu. Dựa vào những thông tin này người sử dụng có thể đánh giá được tiềm năng của dữ liệu đối với các nhu cầu của họ. Bao gồm các thông tin chi tiết như: tên của tập dữ liệu, thời gian thu thập, xuất xứ của tập dữ liệu, vùng địa lý mà tập dữ liệu bao phủ, hiện trạng và vấn đề bảo quản dữ liệu, mục đích và những giới hạn khi sử dụng tập dữ liệu.
- Thông tin chất lượng dữ liệu là nhân tố quan trọng để xác định giá trị đầu tư xây dựng dữ liệu. Chất lượng dữ liệu có tác động tổng thể đến tính hữu ích của dữ liệu, bao gồm các đặc điểm như độ chính xác vị trí, độ chính xác thuộc tính, tính hợp lý, tính toàn vẹn, tính thời gian, nguồn gốc, khả năng truy cập và các chi phí trực tiếp, gián tiếp.
- Thông tin tổ chức dữ liệu không gian cung cấp mô hình dữ liệu được sử dụng để biểu diễn thông tin không gian trong tập dữ liệu. Có 3 loại cấu trúc dữ liệu tổng quát đó là cấu trúc điểm, cấu trúc vector và cấu trúc raster.
- Thông tin tham chiếu không gian mô tả khung tham chiếu dùng cho các tọa độ trong tập dữ liệu bao gồm cả hệ thống tọa độ theo chiều đứng (hệ độ cao, hệ độ sâu) lẫn chiều ngang (hệ tọa độ kinh vĩ, hệ tọa độ địa phương), các điểm mốc tọa độ, độ cao, tỉ lệ, độ phân giải của dữ liệu.
- Thông tin thực thể và thuộc tính dùng để mô tả các thực thể, các thuộc tính, các giá trị thuộc tính và những đặc điểm liên quan được mã hóa trong tập dữ liệu như: tên của thực thể và thuộc tính, loại đối tượng điểm, đường hay vùng, các giá trị thuộc tính theo không gian và thời gian...
- Thông tin phân phối là cơ chế phổ biến tập dữ liệu. Dữ liệu trong các hệ thống thông tin địa lý đòi hỏi tổ chức có sự đầu tư rất lớn về chi phí. Do đó, tổ chức tạo ra dữ liệu cần phải thông tin về tiềm năng của dữ liệu cho những người mua và người sử dụng biết những điều kiện để có thể sử dụng được những tập dữ liệu đó. Thông tin phân phối bao gồm thông tin về định dạng của dữ liệu, hình thức truy cập, công nghệ sử dụng, tính

năng công nghệ, chi phí mà người sử dụng phải bỏ ra để có được nguồn dữ liệu, trách nhiệm của nhà phân phối đối với khách hàng...

- Thông tin tham chiếu Metadata gồm: tên chuẩn, phiên bản của Metadata, ngày tạo ra Metadata, kế hoạch cập nhật, cá nhân/tổ chức chịu trách nhiệm về việc cập nhật dữ liệu, bản quyền truy cập Metadata.
- Thông tin trích dẫn cung cấp cho người sử dụng những thông tin chi tiết về người tạo ra dữ liệu, tạo ra khi nào, ở đâu và phiên bản của tập dữ liệu.
- Thông tin chu kỳ thời gian bao gồm ngày tháng và thời điểm mà dữ liệu được biểu diễn như thời gian bay chụp ảnh vệ tinh để làm cơ sở tạo dữ liệu không gian, thời gian thu thập dữ liệu thuộc tính, thời điểm thống kê dữ liệu gần nhất...
- Thông tin liên lạc cung cấp tên tổ chức, cá nhân chịu trách nhiệm trả lời các câu hỏi về dữ liệu.
- Thông tin lưu trữ và bảo dưỡng cho biết định dạng có thể đọc được tại nơi chứa nguồn dữ liệu đó. Định dạng dữ liệu đó phải ở dạng mở, nghĩa là có thể chuyển đổi được sang các định dạng trên các phần mềm khác. Thông tin lưu trữ gồm các thông tin như: định dạng dữ liệu được lưu trữ, phân cấp lưu trữ, chu kỳ lưu trữ, ngày được lưu trữ và ngày mà dữ liệu có thể được xóa bỏ.

3. VAI TRÒ CỦA METADATA ĐỐI VỚI DỮ LIỆU TRONG CÁC HỆ THỐNG THÔNG TIN ĐỊA LÝ

Thành phần chính của hệ thống thông tin địa lý là dữ liệu. Dữ liệu được dùng để phân tích, truy vấn và hiển thị, không có dữ liệu các phần mềm GIS không thể hoạt động được. Trong khi đó, Metadata giữ vai trò rất quan trọng đối với dữ liệu GIS trong việc lưu trữ và bảo quản những nghiên cứu của tổ chức trên dữ liệu; cung cấp thông tin về các danh mục và ngân hàng dữ liệu; cung cấp thông tin trợ giúp cho việc chuyển đổi và chia sẻ dữ liệu.

Lưu trữ và bảo quản những nghiên cứu của tổ chức vào dữ liệu

Việc xây dựng cơ sở dữ liệu cho một hệ thống thông tin địa lý đòi hỏi phải trải qua một thời gian dài, qua nhiều công đoạn từ việc thu thập, thiết kế cơ sở dữ liệu, tiền xử lý, đăng ký tọa độ, lưới chiếu, số hóa, liên biên, nhập dữ liệu... và nhất là cần rất nhiều người cùng làm việc. Khi những yếu tố con người thay đổi hoặc qua thời gian những thông tin về dữ liệu của tổ chức sẽ bị đánh mất và như thế dữ liệu có thể sẽ mất giá trị. Những người làm việc tiếp tục có thể sẽ không hiểu hết được nội dung, vấn đề xây dựng và sử dụng của dữ liệu dẫn đến công việc sẽ không đạt hiệu quả như mong muốn.

Những mô tả một cách đầy đủ trong Metadata về nội dung và độ chính xác của dữ liệu GIS sẽ đảm bảo lưu trữ lại tất cả những thông tin về tập dữ liệu giúp cho việc sử dụng dữ liệu thích hợp hơn và thấy được sự đầu tư của tổ chức vào dữ liệu. Hơn nữa, Metadata đã trình bày một cách rõ ràng những mục đích và giới hạn của dữ liệu, giúp làm giảm trách nhiệm pháp lý của cơ quan cung cấp dữ liệu nếu có sự lạm dụng dữ liệu để sử dụng trái với mục đích đã đưa ra.

Cung cấp thông tin về các danh mục và ngân hàng dữ liệu

Các ứng dụng của GIS thường đòi hỏi nhiều chủ đề dữ liệu. Một số tổ chức có đủ khả năng để tạo ra tất cả các chuyên đề dữ liệu mà họ cần, thông thường dữ liệu được tạo ra bởi một tổ chức này có thể sử dụng được cho mục đích của tổ chức khác. Metadata cung cấp các

thông tin xuyên suốt toàn bộ các danh mục và ngân hàng dữ liệu, thông qua đó, các tổ chức có thể tìm thấy dữ liệu để sử dụng, tìm thấy đối tác để chia sẻ dữ liệu và tìm kiếm khách hàng cho nguồn dữ liệu của mình.

Cung cấp thông tin trợ giúp cho việc chuyển đổi và chia sẻ dữ liệu

Hầu hết các tổ chức cần dữ liệu nhiều hơn cái mà họ có thể có cũng có khi đó là những dữ liệu ngoài phạm vi quyền hạn thu thập của họ. Các tổ chức không thể thu thập được tất cả các thông tin cơ sở để phục vụ cho các ứng dụng của mình, có một số thông tin tổ chức này không thể thu thập được nhưng các tổ chức khác thì có thể thu thập được. Do đó, cần thiết phải có sự trao đổi, chia sẻ dữ liệu giữa các tổ chức.

Việc chia sẻ dữ liệu không gian có rất nhiều lợi ích, trong đó có 4 lợi ích nổi bật đối với các tổ chức có thực hiện việc chia sẻ dữ liệu không gian đó là: (1) Giảm các chi phí liên quan đến dữ liệu; (2) Tăng cường sự hợp tác trong nội bộ tổ chức và giữa các tổ chức; (3) Tạo ra tính kiên định và tính thống nhất trong tập dữ liệu, (3) Cải tiến chất lượng của quyết định.

Metadata đưa ra một qui trình cung cấp tư liệu đồng nhất cho dữ liệu. Không có Metadata thích hợp, một tổ chức không thể cung cấp tin tức về các thành viên tham gia chia sẻ dữ liệu và thông tin về cấu trúc dữ liệu không gian. ESRI nhận định rằng: "Metadata của dữ liệu không gian rất quan trọng bởi vì chúng lưu giữ các tư liệu về tập dữ liệu đang tồn tại và làm cho việc chia sẻ dữ liệu được dễ dàng. Metadata là một nguồn tài nguyên thiết yếu mà bạn có thể chia sẻ với nhiều người sử dụng tại tất cả các cấp độ của lĩnh vực GIS".

Để thuận lợi cho việc trao đổi dữ liệu, Metadata được yêu cầu để hỗ trợ cho sự phát triển của danh mục cơ sở dữ liệu, cung cấp các mô tả về nội dung, chất lượng, giá cả, các thông số để chuyển đổi dữ liệu và các thông tin có ích khác để dữ liệu dễ dàng được tiếp nhận. Mặc dù việc chia sẻ dữ liệu đến nhiều người sử dụng làm mất cơ hội độc quyền về dữ liệu, nhưng kết quả của sự hợp tác có thể cải thiện hiệu quả công việc, giảm sự dư thừa, giảm bớt chi phí và làm phong phú thêm nguồn tài nguyên dữ liệu cho tổ chức.

Metadata đem lại lợi ích cho cả tổ chức cung cấp dữ liệu và những người sử dụng dữ liệu, việc tạo ra chuẩn Metadata có những thuận lợi:

- Đối với tổ chức cung cấp dữ liệu:

- Metadata tạo điều kiện cho tổ chức lưu lại các thông tin về dữ liệu của chính mình, tránh mất mát thông tin khi có sự thay đổi nhân sự và thời gian.
- Quảng bá nguồn dữ liệu của mình với các tổ chức khác, tìm được nhiều đối tác và khách hàng
- Đưa ra các thông tin chi tiết về nguồn gốc, xuất xứ và vấn đề chất lượng dữ liệu tạo cho người sử dụng sự tin tưởng đối với nguồn dữ liệu.
- Metadata đã trình bày tất cả các vấn đề về dữ liệu một cách đầy đủ và rõ ràng, tổ chức cung cấp dữ liệu không phải mất nhiều thời gian để giải đáp thắc mắc của khách hàng về dữ liệu.

- Đối với người sử dụng dữ liệu:

- Dựa vào Metadata có thể biết được ai, ở đâu đã có những dữ liệu gì, định dạng dữ liệu có phù hợp với công việc của mình không, mất bao nhiêu chi phí để có được nguồn dữ liệu đó.
- Dựa vào Metadata người sử dụng dữ liệu có thể đánh giá được sự phù hợp của tập dữ liệu để sử dụng.

- Tạo nhiều cơ hội để tìm kiếm được nhiều nguồn dữ liệu khác nhau, người sử dụng không phải mất nhiều thời gian và chi phí cho việc xây dựng mới dữ liệu.

4. MÔ HÌNH KHỞI TẠO CHUẨN METADATA CHO DỮ LIỆU GIS

Tiêu chuẩn nội dung Metadata (gọi tắt là chuẩn Metadata) đơn giản là một tập hợp chung của các thuật ngữ và các định nghĩa mô tả dữ liệu GIS. Chuẩn Metadata phát triển để Metadata được tổng hợp trong một khuôn mẫu tương thích. Những tiêu chuẩn có thể giúp các chương trình ứng dụng và công nghệ làm việc chung với nhau. Những tiêu chuẩn khuyến khích hiệu quả và hiệu suất công việc, giúp giảm chi phí, bảo vệ những đầu tư vào dữ liệu chống lại những thay đổi về công nghệ và có thể làm gia tăng độ chính xác, tính đầy đủ và hiện hành của dữ liệu. Việc tạo ra chuẩn Metadata như là tạo thêm một công cụ sử dụng để điều khiển việc phân tích không gian, chuẩn Metadata cung cấp một phương pháp cho những người sử dụng dữ liệu nhận biết: *dữ liệu đang có sẵn cái gì, dữ liệu đáp ứng được những nhu cầu gì, có thể tìm thấy dữ liệu ở đâu và truy cập dữ liệu bằng cách nào.*

Dữ liệu trong hệ thống thông tin địa lý thường là dữ liệu từ nhiều phân hệ khác nhau, việc tập trung tất cả các dữ liệu về một nơi là điều khó khăn và không an toàn. Chuẩn Metadata nhằm giúp cho các phân hệ đưa các thông tin về dữ liệu của mình theo một khuôn mẫu thích hợp để người sử dụng có thể dễ dàng tìm kiếm dữ liệu cũng như các thông tin có liên quan. Khi xây dựng chuẩn Metadata, vấn đề quan trọng là tổ chức sắp xếp nơi lưu trữ thông tin và kiểu lưu trữ để đảm bảo tính đầy đủ và tính chính xác về nội dung của Metadata. Chuẩn Metadata được tổ chức theo mô hình phân cấp của các *yếu tố dữ liệu* và các *yếu tố đa hợp*, đây là các yếu tố xác định nội dung thông tin của Metadata.

Một yếu tố đa hợp là một nhóm các yếu tố dữ liệu và các yếu tố đa hợp khác. Tất cả các yếu tố đa hợp được mô tả trực tiếp bằng các yếu tố dữ liệu hoặc thông qua các yếu tố đa hợp cấp thấp hơn. Hình thức mô tả như sau:

Tên của yếu tố đa hợp – định nghĩa

Kiểu: đa hợp

Tên viết tắt:

Một yếu tố dữ liệu là đơn vị thông tin nhỏ nhất trong tập dữ liệu, là nơi các thông tin tương ứng về dữ liệu sẽ được điền vào. Hình thức mô tả như sau:

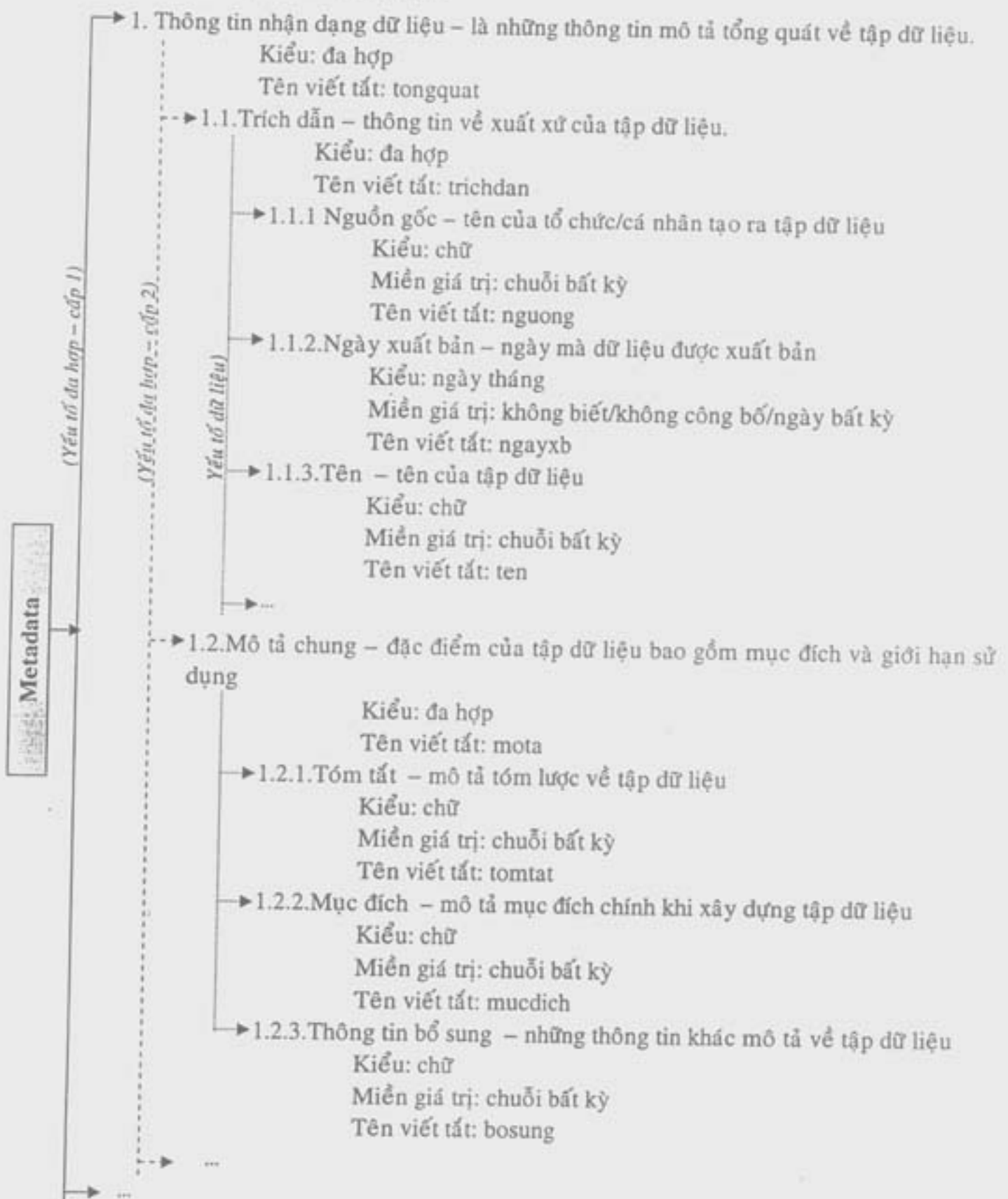
Tên của yếu tố dữ liệu – định nghĩa

Kiểu: số nguyên/số thực/chữ/ngày tháng/giờ

Miền giá trị: giá trị hợp lệ mà có thể được gán vào

Tên viết tắt:

Mô hình minh họa chuẩn Metadata:



Khi xây dựng chuẩn Metadata cho từng tập dữ liệu cụ thể, tùy vào nội dung của tập dữ liệu mà có những thông tin điền vào phù hợp.

5. KẾT LUẬN

Kết quả bước đầu của tiến trình nghiên cứu cũng đã cho thấy được những lợi ích mà Metadata đem lại cho dữ liệu trong các hệ thống thông tin địa lý từ việc lưu trữ, bảo quản các thông tin về dữ liệu, về các danh mục và ngân hàng dữ liệu đến việc cung cấp thông tin trợ giúp cho việc chuyển đổi và chia sẻ dữ liệu và cũng thấy được tính khả thi của việc xây dựng một chuẩn Metadata. Thiết nghĩ, việc nhanh chóng hoàn chỉnh mô hình và phổ biến một chuẩn Metadata cho dữ liệu GIS là một việc làm mang lại nhiều hiệu quả thiết thực. Nếu có thể được, công việc này nên được thực hiện song song với việc nghiên cứu và xây dựng cơ sở dữ liệu GIS, có như thế mới bảo đảm được tính thống nhất và toàn vẹn thông tin trong suốt quá trình xây dựng cơ sở dữ liệu.

TÀI LIỆU THAM KHẢO

- [1] TRẦN VĨNH PHƯỚC. *GIS Một số vấn đề chọn lọc*, NXB Giáo dục, 2001.
- [2] CARMELLE JACQUELINE CÔTÉ. *Content standards for geospatial Metadata: impacts on international spatial data sharing*, dissertation 1998.
- [3] ESRI. *ESRI Profile of the Content Standard for Digital Geospatial Metadata*, 2001.
- [4] ESRI.: *Using ArcCatalog*, ArcInfo 8, Gis by Esri, 1999.
- [5] Federal Geographic Data Committee. *Content Standard for Digital Geospatial Metadata*, Washington, DC, 1998.
- [6] National Mapping Program Technical Instructions, *Standards for Digital Line Graphs*, U.S. Department of the Interior, U.S. Geological Survey, 1999.
- [7] National Research Council. *Distributed Geolibraries*, National Academy, 1999.
- [8] STAN ARONOFF. *Geographic Information Systems: A management Perspective*, WDL Publications, 1993.