

# MỘT SỐ CHUẨN MÔ TẢ DỮ LIỆU

## NHẪM LIÊN KẾT VÀ CHIA SẺ TÀI NGUYÊN THÔNG TIN GIỮA CÁC THƯ VIỆN

**Đinh Thuý Quỳnh**

Khoa Thư viện - Thông tin, Đại học Văn hoá Hà Nội

Tóm tắt:

Ngày nay, không một thư viện nào có đủ kinh phí để có thể bổ sung đầy đủ nguồn lực thông tin phục vụ cho nhu cầu ngày càng đa dạng và phong phú của người dùng tin. Vì vậy, việc liên kết và chia sẻ tài nguyên thông tin giữa các thư viện không chỉ trong phạm vi quốc gia mà cả quốc tế là điều tất yếu. Để làm được điều đó cùng với sự hỗ trợ tối đa của thiết bị khoa học công nghệ thì đòi hỏi các thư viện cần phải sử dụng chung các chuẩn nghiệp vụ, trong đó có chuẩn mô tả dữ liệu. Bài viết giới thiệu một số chuẩn mô tả cho cả dữ liệu thư mục và dữ liệu số để các thư viện có thể tham khảo và trên cơ sở căn cứ vào nhu cầu sử dụng cũng như quy mô mà các thư viện sẽ đưa ra những quyết định trong việc lựa chọn các chuẩn mô tả phù hợp.

**Từ khoá:** Thư viện, mô tả tài liệu, chuẩn quốc tế.

Abstract:

Nowadays, no single library has enough budgets to acquire all resources it needs to serve its clients. Therefore, the collaboration and sharing of information resources between libraries at local and international levels are unavoidable. In order to make it happen, in addition to the support of ICT and modern equipment, libraries need to use standardized tools including data description standards. This paper introduces some standards for description of bibliographic and digital data, which supports libraries in selecting the most relevant standards for their operation.

**Keywords:** Library, description of works, international standard.

Chuẩn mô tả dữ liệu là việc định ra một trật tự nhất định cho các yếu tố mô tả và cách thức quy ước ghi ký hiệu dấu cho từng yếu tố mô tả đó. Chuẩn mô tả dữ liệu quốc tế được đưa ra nhằm thực hiện việc thống nhất mô tả tài liệu trong các thư viện trên toàn thế giới. Nó đẩy mạnh việc trao đổi thông tin giữa các nước với nhau, khắc phục những khó khăn do sự bất đồng ngôn ngữ, đồng thời thiết lập được một hệ thống thông tin giữa các nước, thực hiện tự động hoá trong công tác biên mục. Hiện nay, có khá nhiều chuẩn mô tả cho cả dữ liệu dạng thư mục và dạng số. Sau đây, chúng ta cùng điểm lại một số chuẩn mô tả dữ liệu quốc tế đã và đang được sử dụng phổ biến trong nhiều thư viện lớn trong nước và trên thế giới:

### 1. Chuẩn mô tả dữ liệu thư mục

#### 1.1. Quy tắc mô tả thư mục theo tiêu chuẩn quốc tế ISBD (International Standard Bibliographic Description)

Quy tắc mô tả thư mục theo tiêu chuẩn quốc tế ISBD được Liên đoàn Quốc tế các Hội và cơ quan Thư viện (IFLA - International Federation of Library Association an Institution) thông qua nghị quyết về biên soạn vào năm 1969 tại Hội nghị chuyên gia biên mục được tổ chức ở Đan Mạch.

ISBD chỉ đề cập tới các yếu tố được sắp xếp dưới tiêu đề mô tả. Với nội dung đó, ISBD sẽ phản ánh đầy đủ mọi thông tin cần thiết về tài liệu kể cả các chi tiết về tác giả giúp việc xác định tài liệu được dễ dàng.

Tiêu đề mô tả trong ISBD không mang tính bắt buộc. Tuy nhiên, để phù hợp với yêu cầu sắp xếp tài liệu trong mục lục chữ cái cũng như đáp ứng yêu cầu tra cứu của bạn đọc, tại thư viện của một số nước, tiêu đề mô tả đã trở thành yếu tố bắt buộc đối với việc mô tả tài liệu. Những thông tin cần thiết để xác định tài liệu được sắp xếp thành 8 vùng mô tả. Trong mỗi vùng có nhiều yếu tố khác nhau, có những yếu tố bắt buộc và những yếu tố không bắt buộc. Điều này thể hiện tính mềm dẻo của ISBD phù hợp với công tác mô tả ở tất cả các loại hình thư viện từ nhỏ đến lớn.

Các vùng mô tả gồm:

1. Vùng nhan đề và các thông tin về trách nhiệm (Nhan đề chính, nhan đề song song, thông tin liên quan đến nhan đề, các thông tin về trách nhiệm)
2. Vùng lần xuất bản và thông tin về trách nhiệm có liên quan đến lần xuất bản.
3. Vùng thông tin đặc thù (cho xuất bản phẩm nhiều kỳ và tư liệu chuyên dụng).
4. Vùng địa chỉ xuất bản (nơi xuất bản, nhà xuất bản, năm xuất bản).
5. Vùng đặc trưng số lượng (khối lượng: số trang, tờ, cột, minh họa hoặc khổ sách, tài liệu kèm theo).
6. Vùng từng thư (Nhan đề chính của từng thư, nhan đề song song của từng thư, các thông tin bổ sung cho nhan đề từng thư, thông tin về trách nhiệm của từng thư, số tập).
7. Vùng phụ chú.
8. Vùng ISBN: kiểu đóng, giá tiền.

Hệ thống các ký hiệu: Hệ thống các ký hiệu dấu của ISBD để nhận dạng các vùng và các yếu tố nhằm tạo điều kiện để nhận biết chúng trong phiếu mô tả một cách dễ dàng. Mỗi dấu được quy định đặt trước hoặc sau một vùng hoặc một yếu tố. Việc dùng các ký hiệu dấu này không ảnh hưởng tới các dấu chính tả thông thường.

Có hai loại dấu ký hiệu:

- Ký hiệu dùng chung cho các vùng như: dấu chấm gạch ngang ( . - ); dấu ngoặc đơn, dấu ba

chấm...

- Ký hiệu dùng riêng cho từng yếu tố như: dấu chấm, dấu phẩy, dấu chấm phẩy...

Có thể nói rằng áp dụng ISBD để mô tả tài liệu cũng không ảnh hưởng gì đến vấn đề tổ chức kỹ thuật của thư viện, không đảo lộn hệ thống các mục lục đã có sẵn trong thư viện, bởi vì ISBD chỉ quy định lại trật tự sắp xếp các yếu tố và cho mỗi yếu tố một ký hiệu dấu nhất định. Còn tiêu đề mô tả áp dụng trong mục lục vẫn giữ nguyên theo các quy định mô tả cũ. Tiêu chuẩn này nhanh chóng được áp dụng vào các thư viện quốc gia, nơi phải thực hiện mô tả đầy đủ cho các thư mục quốc gia và ngày càng được sử dụng rộng rãi ở nhiều nước. Ở Việt Nam, chuẩn ISBD đã được áp dụng rộng rãi trong các thư viện từ năm 1985 cho đến nay.

Trải qua một quá trình chỉnh sửa và bổ sung, một loạt quy tắc mô tả theo ISBD chuyên dụng ra đời như:

- ISBD (M - Monographic): Mô tả ấn bản phẩm chuyên khảo.
- ISBD (CM - Cartographic Materials): Mô tả tư liệu bản đồ.
- ISBD (NBM - Non book Materials): Mô tả tư liệu không phải là sách báo.
- ISBD (S - Serial): Mô tả ấn bản phẩm định kỳ.
- ISBD (CF - Computer files): Mô tả tệp tin hay cơ sở dữ liệu.
- ISBD (A - Antiquarial): Mô tả sách cổ.
- ISBD (PM - Printed music): Mô tả nhạc phẩm in.
- ISBD (G - General): Mô tả chung.
- ISBD (CP - Component Parts): Mô tả trích.

Hiện nay ISBD (CF) đã chuyển thành - ISBD (ER - Electronic Resource) để bao quát thêm nguồn tin điện tử.

### **1.2. Quy tắc biên mục Anh Mỹ - AACR2 (Anglo-American Cataloguing Rules)**

Năm 1967, người Mỹ và người Anh đã hợp tác

biên soạn và xuất bản một bộ quy tắc chung gọi tắt theo tiếng Anh là AACR dành cho các thư viện nghiên cứu lớn. AACR đưa ra các quy tắc mô tả đối với các tư liệu không phải sách báo. Quy tắc này nhấn mạnh các thông tin trên trang nhan đề, chú trọng lập tiêu đề theo các loại tên người hơn là thể loại các tác phẩm, không nhấn mạnh tới việc biên soạn các quy tắc riêng theo từng loại hình tư liệu, phân biệt mô tả chính và mô tả bổ sung, phân biệt hai hình thức trình bày mô tả: theo tác giả và theo nhan đề.

AACR còn đề cập đến hình thức mô tả với tiêu đề tác giả tập thể, là tên gọi cho các cơ quan, tổ chức và những quy định đối với việc đưa tên gọi của các cơ quan cấp trên vào tiêu đề. Song AACR còn nhiều nhược điểm khi có những quy định bổ sung và ngoại lệ mang tính chắp vá.

Năm 1974, đại diện của các hội thư viện và các thư viện quốc gia Anh, Mỹ, Canada đã họp và soạn thảo một bộ quy tắc hoàn toàn mới - AACR2, khắc phục tình trạng chắp vá và nói trên, tạo điều kiện thuận lợi cho người sử dụng mục lục nhanh chóng tìm được tư liệu và cải thiện hợp tác quốc tế trong lĩnh vực biên mục.

Bộ quy tắc AACR2 gồm 19 chương, chia làm 2 phần:

Phần I, từ chương 1 đến chương 13 là phần Mô tả thư mục. Phần này quy định cách mô tả các loại hình tài liệu khác nhau và dựa trên quy định của ISBD (Quy tắc mô tả thư mục theo tiêu chuẩn Quốc tế).

Phần II, từ chương 21 đến chương 26 là phần Tiêu đề, tên sách thống nhất và tham chiếu.

AACR2 đặt nền tảng cho sự hợp tác biên mục trên phạm vi quốc gia và quốc tế, cải tiến các dịch vụ thư mục và tiết kiệm được giá thành. Do cung cấp một hệ thống mô tả chuẩn mực thống nhất cho tất cả các loại hình tư liệu nên quy tắc này đã tạo khả năng thực hiện mục lục tích hợp đa phương tiện. Nhìn chung không có gì khác biệt nhiều giữa AACR2 và ISBD về 8 vùng mô tả và dấu phân cách. Tuy nhiên, AACR2 tạo nhiều điểm truy cập và quy định khá chi tiết trong lập tiêu đề cũng như trong một số yếu tố mô tả. Bên cạnh đó, đây là bộ quy tắc chuẩn quốc tế đã được đa

số các thư viện trên thế giới dùng. Do vậy, một thư viện khi muốn hội nhập với cộng đồng thư viện thế giới thì nên sử dụng bộ quy tắc biên mục AACR2.

### **1.3. Quy tắc thống nhất mô tả ấn phẩm cho mục lục thư viện của Liên Xô**

Quy tắc thống nhất mô tả ấn phẩm cho mục lục thư viện được biên soạn dựa theo truyền thống mô tả Anh - Mỹ và những kinh nghiệm biên mục của các thư viện lớn ở Liên xô cũ. Đặc điểm của bộ quy tắc này là bao quát được mọi loại hình tư liệu được xuất bản thời kỳ bấy giờ và chú trọng tới việc làm rõ nội dung tư liệu, xác định rõ phạm vi sử dụng mô tả tác giả tập thể.

Bộ quy tắc gồm:

- Phần I:

o Tập 1: Mô tả sách

o Tập 2: Tổ chức mục lục chữ cái sách

- Phần II: Mô tả xuất bản phẩm định kỳ

- Phần III: Mô tả bản đồ

- Phần IV: Mô tả nhạc phẩm

- Phần V: Mô tả ấn phẩm đồ hoạ

- Phần VI: Các loại tư liệu chuyên dụng kỹ thuật

- Phần VII: Mô tả tư liệu ghi âm và nghe nhìn

- Phần VIII: Mô tả tư liệu rời rạc với nội dung chính trị, xã hội.

Trong quá trình áp dụng bộ quy tắc này đã dần được bổ sung, sửa đổi và hoàn thiện trên tinh thần của ISBD.

Các quy tắc này cũng là cơ sở để biên soạn và ban hành hàng loạt tiêu chuẩn Liên Xô như:

- GOST 7.1-69: Mô tả ấn phẩm dùng cho các xuất bản phẩm thư mục và thông tin

- GOST 7.2-69: Mô tả xuất bản phẩm định kỳ cho các mục lục

- GOST 7.3-69: Mô tả các loại tư liệu chuyên dụng kỹ thuật cho các mục lục... là công cụ pháp lý để thống nhất hoá mô tả ấn phẩm đối với các

thư viện, cơ quan thông tin, các nhà xuất bản và phát hành trong toàn liên bang.

Quy tắc này cũng đã có ảnh hưởng lớn đến thực tiễn biên mục của Việt Nam trong hàng thập kỷ từ đầu những năm 1960 đến những năm 80 của thế kỷ 20 khi ảnh hưởng của ISBD bắt đầu xâm nhập vào nước ta.

## 2. Chuẩn mô tả dữ liệu số

### 2.1. Bộ yếu tố siêu dữ liệu Dublin Core

Dublin Core là một trong những lược đồ yếu tố siêu dữ liệu phổ biến và được nhiều người biết đến. Bộ yếu tố này được hình thành lần đầu tiên vào năm 1995 bởi sáng kiến Yếu tố Siêu dữ liệu Dublin Core (Dublin Core Metadata Element Initiative). Tập hợp yếu tố siêu dữ liệu này được gọi là "cốt lõi" (core) vì nó được thiết kế đơn giản và chỉ bao gồm 15 yếu tố mô tả cốt lõi nhất (trong khi MARC21 có hơn 200 trường và rất nhiều trường con). Tháng 9/2001, bộ yếu tố siêu dữ liệu Dublin Core được ban hành thành tiêu chuẩn Mỹ, gọi là tiêu chuẩn "The Dublin Core Metadata Element Set" ANSI/NISO Z39.85-2001.

Siêu dữ liệu Dublin Core bao gồm 15 yếu tố sau:

- **Nhan đề (Title):** Nhan đề của tài liệu.

- **Tác giả (Creator):** Tác giả của tài liệu, bao gồm cả tác giả cá nhân và tác giả tập thể.

- **Chủ đề (Subject):** Chủ đề tài liệu đề cập dùng để phân loại tài liệu. Có thể thể hiện bằng từ, cụm từ/ (Khung chủ đề), hoặc chỉ số phân loại/ (Khung phân loại).

- **Tóm tắt (Description):** Tóm tắt, mô tả nội dung tài liệu. Có thể bao gồm tóm tắt, chú thích, mục lục, đoạn văn bản để làm rõ nội dung...

- **Nhà xuất bản (Publisher):** Nhà xuất bản, nơi ban hành tài liệu có thể là tên cá nhân, tên cơ quan, tổ chức, dịch vụ...

- **Tác giả phụ (Contributor):** Tên những người cùng tham gia cộng tác đóng góp vào nội dung tài liệu, có thể là cá nhân, tổ chức...

- **Ngày tháng (Date):** Ngày, tháng ban hành tài liệu. Có thể dùng chuẩn ISO 8601.

- **Loại (kiểu) (Type):** Mô tả bản chất của tài liệu. Dùng các thuật ngữ mô tả phạm trù kiểu: trang chú, bài báo, báo cáo, tử điển...

- **Khổ mẫu (Format):** Mô tả sự trình bày vật lý của tài liệu, có thể bao gồm: vật mang tin, kích cỡ độ dài, kiểu dữ liệu (.doc, .html, .jpg, xls, phần mềm...).

- **Định danh (Identifier):** Các thông tin về định danh tài liệu, các nguồn tham chiếu đến, hoặc chuỗi ký tự để định vị tài nguyên: URL, URN (Uniform Resource Name), ISBN, ISSN, SICI (Serial Item & Contribution Identifier).

- **Nguồn (Resource):** Các thông tin về xuất xứ của tài liệu, tham chiếu đến nguồn mà tài liệu hiện mô tả được trích ra/ tạo ra, nguồn cũng có thể là: đường dẫn (URL), URN, ISBN, ISSN.

- **Ngôn ngữ (Language):** Các thông tin về ngôn ngữ, mô tả ngôn ngữ chính của tài liệu: Có thể sử dụng chuẩn ISO 639 để mô tả ngôn ngữ cho tài liệu.

- **Liên kết (Relation):** Mô tả các thông tin liên quan đến tài liệu khác, có thể dùng đường dẫn (URL), URN, ISBN, ISSN.

- **Diện bao quát (Coverage):** Các thông tin liên quan đến phạm vi, quy mô hoặc mức độ bao quát của tài liệu. Phạm vi đó có thể là địa điểm, không gian hoặc thời gian, toạ độ...

- **Bản quyền (Right):** Các thông tin liên quan đến bản quyền của tài liệu.

Mục đích của Dublin Core là cung cấp một bộ phần tử dữ liệu đơn giản cho việc mô tả tài liệu và các đối tượng khác trên Internet. Tức là mọi tài liệu trên Internet sẽ được mô tả thư tịch riêng và chúng sẽ có những phần tử dữ liệu được mã hoá cho các phần thông tin như tác giả, nhan đề, ngày tháng. Theo tiêu chuẩn này, một tài liệu có thể được tìm theo tên tác giả hoặc nhan đề. Trên Internet, Dublin Core được sử dụng ngày càng nhiều, Dublin Core đã trở thành một siêu dữ liệu mô tả tài liệu cho nhiều ứng dụng dựa trên web...

Dublin Core cung cấp một bộ phần tử chính có thể dễ dàng nhúng vào nhiều ứng dụng web khác nhau. Dublin Core có thể sử dụng bởi nhiều cộng

đồng văn hoá khác nhau. Do đó, nó đã làm tăng tính mở giữa các cộng đồng khi muốn trao đổi và chia sẻ tài nguyên cho nhau.

## **2.2. Khổ mẫu MARC (MACHine Readable Cataloguing)**

MARC là khổ mẫu cho phép máy tính lưu trữ và truy xuất thông tin. Mục lục tin học hoá bao gồm các biểu ghi theo khổ mẫu MARC. Có nghĩa là người biên mục cần mã hoá định nhãn thông tin trong biểu ghi.

Khổ mẫu MARC là một cấu trúc dành riêng cho các dữ liệu thư mục đưa vào máy tính. Nét độc đáo của nó là đưa ra một phương pháp mã hoá rộng rãi dựa trên cơ sở phân tích sâu sắc những yếu tố thư mục. Một biểu ghi thư mục được thể hiện trong khổ mẫu gồm 3 phần: cấu trúc, nội dung và mã hoá. Cấu trúc và mã hoá là đặc điểm của khổ mẫu nói riêng, còn nội dung thì được hình thành từ các yếu tố dữ liệu thư mục như đã được xác định trong các quy tắc biên mục. Biểu ghi là tập hợp của các trường, mỗi trường chứa đựng một đơn vị thông tin trong các biểu ghi. Một trường có thể chia thành nhiều trường con. Các nhãn trường là các mã số gồm 3 chữ số dùng để nhận dạng trường, mỗi trường kết thúc bằng một dấu kết thúc trường. Trước mỗi trường con đều có dấu phân định.

Việc tiếp cận tới các trường có thể được thực hiện trực tiếp nhờ một danh sách ghi các nhãn trường, độ dài và địa chỉ. Các nhãn trường có 3 chữ số cho phép tạo ra 999 vùng thoả mãn mọi nhu cầu nhận dạng tư liệu. Các nhãn đó xác định dữ liệu theo 2 mức độ: về bản chất (như tên người) và về chức năng (như tác giả chính). Mỗi trường có 2 chỉ số để nói về các trường hợp xử lý đặc biệt (như chỉ rõ thêm nội dung của trường hay quan hệ giữa trường này với trường khác...). Người ta còn sử dụng mã hoá trong những trường cố định nhằm nêu rõ hình thức và nội dung tư liệu (ngôn ngữ, nước, luận án...) và tạo điều kiện sắp xếp dễ dàng. Các trường con giúp phân biệt các yếu tố chính với các yếu tố phụ (như tên riêng của các tác giả, phụ đề...). Cấu trúc biểu ghi này tạo ra nhiều khả năng sắp xếp, chọn lọc, đánh chỉ số, in ấn, tìm tin và hiệu đính.

Cấu trúc của khổ mẫu MARC là cấu trúc biểu ghi, bao gồm các phần sau:

- **Đầu biểu (Leader):** Là một trường dữ liệu đặc biệt có độ dài cố định chứa các thông tin về quá trình xử lý biểu ghi.

- **Danh mục (Directory):** Là phần tiếp sau ngay phần thông tin đầu biểu, là một chỉ dẫn về các trường dữ liệu có trong biểu ghi.

- **Các trường dữ liệu thư mục (Bibliographic Data):** Là những trường dữ liệu có trong biểu ghi và chứa các dữ liệu mô tả, có thể có độ dài biến đổi (variable fields) hoặc cố định (fixed field).

Với cấu trúc này, khổ mẫu MARC đã tạo ra nhiều khả năng cho máy tính lựa chọn và sắp xếp các dữ liệu thư mục:

- Cho phép người sử dụng dễ dàng truy cập tới các biểu ghi.

- In ra các thông báo sách mới, các ấn phẩm thư mục, các mục lục dưới dạng thức khác nhau, các nhãn trên gáy sách.

- Trao đổi dữ liệu thư mục với các thư viện khác trong nước và trên thế giới.

Khổ mẫu MARC có ý nghĩa quan trọng trong biên mục tự động. Vì vậy, các phần mềm tư liệu và phần mềm quản trị thư viện cần phải được xây dựng tuân theo các chuẩn của khổ mẫu MARC.

## **2.3. Tiêu chuẩn về mã hoá và truyền siêu dữ liệu METS (Metadata Encoding and Transmission Standard)**

METS là một tiêu chuẩn mới được thiết kế để mã hoá các loại siêu dữ liệu cho việc mô tả hoàn chỉnh một đối tượng trong thư viện số. Đó có thể là các văn bản, hình ảnh tĩnh, video, âm thanh hay các loại tài liệu phức tạp khác như môi trường VRML (Virtual Reality Modeling Language - Ngôn ngữ mô phỏng thực tại ảo). Cho đến gần đây, chưa có phương pháp chuẩn hoá nào cho việc mã hoá các siêu dữ liệu cho các đối tượng trong thư viện số. Vì thế các dự án thư viện số thường có xu hướng hoạt động theo cách của riêng và họ thường sử dụng các phần mềm và các định dạng dữ liệu mà các đội dự án của họ đã quen làm. Khi

số lượng các dự án này tăng lên thì sự cần thiết phải có một tiêu chuẩn cho siêu dữ liệu trở nên cấp thiết hơn. Và chính nhu cầu này đã dẫn đến sự hình thành nên METS. Liên đoàn thư viện số đã đưa ra 3 loại siêu dữ liệu chính cần thiết cho việc mô tả các đối tượng của thư viện số là Siêu dữ liệu mô tả (Descriptive Metadata), Siêu dữ liệu quản trị (Administrative Metadata), Siêu dữ liệu cấu trúc (Structural Metadata).

Chuẩn này là sự tổng hợp tất cả các loại hình siêu dữ liệu: mô tả, quản trị và cấu trúc. METS được thiết kế một cách chi tiết cho việc mô tả dữ liệu ảnh và văn bản trong lược đồ XML Schema (đã được thảo luận ở trên). Các tài liệu METS có các phần và yếu tố liên quan hầu hết đến đối tượng số, ngoài ra nó có khả năng mô tả các đối tượng xác định bao gồm việc tham chiếu đến siêu dữ liệu mở rộng theo các định dạng được chuẩn hoá.

Cấu trúc của METS: Một tài liệu METS bao gồm 7 thành phần chính sau:

- **Đầu mục METS (METS Header):** Bao gồm siêu dữ liệu về bản thân tài liệu của METS như: tác giả, người tạo, ngày tháng, sản phẩm, trạng thái...;

- **Siêu dữ liệu mô tả (Descriptive Metadata):** Có một hoặc nhiều trường hợp của siêu dữ liệu mô tả về đối tượng thư viện số, có thể xem xét siêu dữ liệu ngoài tài liệu METS hoặc bên trong hoặc cả hai;

- **Siêu dữ liệu quản trị (Administrative Metadata):** Cung cấp thông tin về các file được tạo, lưu trữ, quyền sở hữu trí tuệ, thông tin tài nguyên như thế nào... tất cả có thể nằm trong hoặc ngoài tài liệu;

- **File tóm tắt (File Inventory):** Danh sách tất cả các file gồm nội dung thực sự cho đối tượng thư viện số;

- **Sơ đồ cấu trúc (Structural Map):** Là thành phần chính của tài liệu METS. Cấu trúc này tóm tắt cấu trúc phân cấp cho đối tượng và liên kết các thành phần của cấu trúc tới các file nội dung;

- **Liên kết cấu trúc (Structural Links):** Cho phép người tạo nội dung có thể ghi lại các liên kết

giữa các nút trong cấu trúc logic trong sơ đồ cấu trúc;

- **Trạng thái (Behaviors):** Là một bản ghi các hoạt động phần mềm để truy nhập đối tượng METS hoặc bất kỳ thành phần nào của nó. Mỗi hoạt động này có một giao diện mô tả tóm tắt các hoạt động cho mỗi phiên hoạt động nhất định.

Để có thể có được toàn bộ lợi ích của việc trao đổi siêu dữ liệu thì cần phải có 2 loại chuẩn hoá: chuẩn hoá siêu dữ liệu của chính vật chứa (giống như định dạng MARC trong lĩnh vực thư viện truyền thống) và việc chuẩn hoá nội dung được chứa trong vật chứa đó (cũng giống như quy tắc biên mục và các hoạt động trong thư viện truyền thống). Định dạng METS cung cấp định dạng thứ nhất, một tiêu chuẩn nhưng linh hoạt dùng để lưu trữ các siêu dữ liệu đa dạng kết hợp với đối tượng số đó trong một dạng mà có thể chia sẻ, tra cứu chéo, trao đổi, tìm kiếm và dùng cho các mục đích duyệt qua và trình diễn.

#### 2.4. Tiêu chuẩn mô tả đối tượng siêu dữ liệu MODS (Metadata Object Description Schema)

Trong một môi trường kết nối mạng mà ở đó siêu dữ liệu mô tả được chuyển tiếp qua nhiều hệ thống và có thể gắn liền vào trong hoặc với nhiều loại siêu dữ liệu khác, thì biểu ghi MARC phù hợp cho mục đích này. Tuy nhiên, MARC không phải là một biểu ghi XML. Bên cạnh đó, biểu ghi MARC lớn và chi tiết hơn mọi hệ thống có thể cần, việc sử dụng các thẻ số và mã trường con làm cho mọi người khó hiểu nếu không được đào tạo đầy đủ. Bởi vậy, người ta đã cần một phiên bản MARC nhẹ nhàng và đơn giản hơn mà có thể chấp nhận những phần tử dữ liệu chính từ một biểu ghi MARC và chuyển chúng vào một khổ mẫu XML dễ dàng hơn. Vì vậy tiêu chuẩn MODS đã ra đời để giải quyết vấn đề này.

MODS sử dụng những thẻ mà người ta dễ dàng hiểu để thay thế cho những thẻ bằng 3 chữ số và mã trường con của MARC (ví dụ: "Nhan đề" ("title") thay thế cho "245"). Nó bỏ qua phần lớn các phần tử dữ liệu của trường cố định, ngoại trừ những mã khổ mẫu có tính vật lý (của mã 007) và nhiều mã dành cho thể loại (của mã 008). Nó cũng giới thiệu một số cách thực hành hiệu quả

và mới mẻ hơn. MODS định nghĩa một cấu trúc gọi là "Tên" ("Name") đại diện nhiều trường bao gồm cả trường con để chứa đựng thông tin tên cá nhân, công ty và cho tên một hội nghị. Cấu trúc này có thể được sử dụng bất kỳ khi nào một tên xuất hiện, nó có thể là mục dữ liệu chính, phụ hoặc chủ đề.

MODS cung cấp 19 thành tố mức đỉnh cho việc mô tả các đối tượng và hơn 64 thành tố con mức dưới. Các thành tố này được dựa trên các khía cạnh thư mục như tiêu đề, tên của người tạo và người xây dựng, số lượng các phân loại và chủ đề, ngoài ra cũng gồm các thành phần thuộc lĩnh vực mô tả vật lý, thông tin về sự giới hạn truy cập, và các thể loại. MODS cũng bao gồm các cơ chế cho phép mở rộng tập các thành tố của nó thông qua việc trao đổi thông tin với các tài liệu XML khác dựa trên cấu trúc của nó; MODS cũng cho phép lưu lại các đối tượng liên quan để có thể tương tác theo cách đó. Ngoài ra, MODS còn có thể dễ dàng điều khiển tính xác thực và mô tả chi tiết theo mức mà nó cung cấp.

MODS được xây dựng khi mà các xu hướng khác nhau về mô tả siêu dữ liệu không có sự thống nhất: nó làm tương thích các vấn đề của Dublin Core bằng cách cung cấp một tập các thành tố đa dạng và cho phép mở rộng thêm khi cần. Cũng như một ứng dụng XML, MODS có khả năng hoạt động độc lập và không ràng buộc với bất kỳ gói phần mềm nào. Các công cụ được viết có thể chuyển đổi qua lại sang các bản ghi MARC XML.

MODS được sử dụng cho một lượng lớn các dạng tài liệu khác nhau từ sách tới các tài liệu đa phương tiện và được lựa chọn trong phần lớn các dự án thư viện số. MODS còn được sử dụng rộng rãi trong các phần chính trong thế giới thư viện số. Khi sử dụng chung với các chuẩn khác như METS, nó làm tăng cường khả năng liên kết đầy đủ và chiến lược tích hợp siêu dữ liệu làm tăng truy cập các tài liệu số trên toàn cầu.

Tuy nhiên MODS cũng đang trong quá trình phát triển và hoàn thiện, nên những hạn chế tiềm ẩn là có thể xảy ra.

Các thẻ trong tài liệu MODS:

**Thông tin nhan đề (Title Info):** Là một yếu tố mức đỉnh bắt buộc, yếu tố này gồm 5 yếu tố con, một trong số đó là **nhan đề**, đây là thành phần bắt buộc. Nhan đề được sử dụng để ghi lại phần bắt buộc. Nhan đề được sử dụng để ghi lại tiêu đề chính cho tài liệu và các yếu tố ngang cấp của nó có thể được ghi vào cùng các thành phần đó như tiêu đề phụ, hoặc là một phần nào đó của đối tượng để ghi chú tiêu đề và các chi tiết khác của tiêu đề.

**Tên (Name):** Yếu tố này tương đương với trường *Tác giả* và *Tác giả phụ* trong Dublin Core, **tên** được sử dụng để lưu tên người hoặc các tổ chức có khả năng tạo nội dung của tài liệu, hoặc ghi lại chính những người tạo nên tài liệu (như người mô tả hoặc người in).

**Thông tin gốc (Origin Info):** Là một yếu tố gộp khác, mang thông tin tổng hợp có nguồn gốc hoặc phổ biến của tài liệu. Các yếu tố con lưu lại ngày tạo hoặc cập nhật tài liệu (trong trường hợp các tài liệu không được phổ biến hoặc là các bản thảo), hoặc lưu vết của các tài liệu gốc.

**Mô tả vật lý (Physical Description):** Là một yếu tố gộp bao gồm các yếu tố con khác nhau cho phép mô tả một cách cơ bản các tính chất vật lý của đối tượng. Các tính chất này chỉ liên quan đến các tài nguyên điện tử: chúng bao gồm Internet MediaType, là thành phần ghi lại định dạng dữ liệu được mô tả (thường có định dạng theo chuẩn MIME như "text/html"), Reformatting Quality, chỉ định về chất lượng (về độ phân giải và số màu) khi tài liệu được scan, Digital Origin (nguồn gốc số) ghi lại xem đối tượng nào được số hoá hoặc đối tượng nào được định dạng lại theo một thiết bị khác. Hầu hết các phương tiện truyền thống có thể lưu lại thông tin tương đối theo các phương thức hạn chế, chủ yếu có yếu tố mở rộng (extent) ghi lại số trang, và yếu tố ghi chú (note) ghi lại các thông tin không cấu trúc về các đặc tính vật lý của đối tượng.

**Chủ đề (Subject):** Là một yếu tố gộp được sử dụng để mô tả đề mục của tài liệu theo các quy tắc phân loại nào đó. Yếu tố con Địa lý phân cấp (Hierarchical Geographic) có thể dùng để định nghĩa sự phân cấp của geographic terms (điều kiện địa lý), cho phép duyệt từ nhiều nơi (như phạm vi lục địa), từ nhiều địa điểm xác định

(như thành phố). Yếu tố con khác là bản đồ (cartographics), cho phép lưu lại chi tiết theo tọa độ địa lý và theo phạm vi trong bản đồ.

**Tài liệu liên quan (Related Item):** Đây là một yếu tố rất hữu ích trong ngữ cảnh của bộ sưu tập các tài liệu có mối liên hệ với nhau (như bộ sưu tập các bài báo và tuần san được số hoá), tài liệu liên quan cho phép các tài liệu liên quan có thể được nhúng theo các thẻ có liên quan trong các tài liệu MODS đầy đủ. Thuộc tính **type** xác định hình thức của mối quan hệ. Yếu tố này mang cùng chức năng như **Relation** (liên kết) trong Dublin Core, nhưng có khả năng mềm dẻo hơn trong cách sử dụng.

Mặc dù MODS cung cấp tập các yếu tố mở rộng hơn Dublin Core, nhưng nó vẫn không thể làm thoả mãn tất cả các yêu cầu siêu dữ liệu cho đối tượng. Trong trường hợp đó, nó cung cấp một cơ chế để dễ dàng mở rộng tập các yếu tố bằng cách cho phép ghi lại siêu dữ liệu trong các lược đồ có khả năng thay đổi được gắn vào trong tài liệu MODS record.

Tóm lại, MODS cung cấp một tập các điều khoản phong phú và có độ linh hoạt cao cho phép tương tác với các loại hình khác nhau.

#### **Một số yếu tố khác của MODS**

Các yếu tố MODS mức đỉnh còn lại là các yếu tố mang tính tự định nghĩa:

**Thuộc tính của nguồn (Type of Resource):** Là kiểu của đối tượng được ghi như: văn bản, đa phương tiện, bản đồ. Các quy định sử dụng được lấy từ danh sách riêng.

**Thể loại (Genre):** Là một quy định đặc biệt hơn type of Resource, yếu tố này cho phép quy định phân loại chi tiết để mô tả cho tài liệu.

**Ngôn ngữ (Language):** Ghi lại ngôn ngữ mô tả của tài liệu.

**Tóm tắt (Abstract):** Mô tả nội dung tóm tắt của tài liệu hoặc một liên kết tới một định dạng tương tự như vậy.

**Mục lục (Table of Contents):** Liệt kê nội dung của tài liệu đã được ghi lại một cách rõ ràng hoặc được cung cấp như một liên kết tới danh

sách đó.

**Người sử dụng mục tiêu (Target Audience):** Quy định các đối tượng người sử dụng (ví dụ như vị thành niên, hoặc thanh niên) cho tài liệu.

**Ghi chú (Note):** Là yếu tố ghi lại các thông tin cần lưu ý.

**Phân loại (Classification):** Chỉ số phân loại tài nguyên theo một lược đồ đã được thông qua như tiêu đề trong Thư viện Quốc hội Mỹ hoặc phân loại thập phân Dewey.

**Định danh (Identifier):** Số hoặc mã duy nhất phù hợp với lược đồ đã được thông qua như số ISBN hoặc ISSN.

**Vị trí (Location):** Ghi lại vị trí vật lý của tài liệu, bao gồm kho lưu trữ và chỉ số giá của tài liệu.

**Hạn chế truy cập (AccessRestriction):** Thông tin về việc truy cập tài liệu bị hạn chế thế nào, bao gồm thông tin về bản quyền.

**Thông tin biểu ghi (RecordInfo):** Một yếu tố gộp về thông tin tạo bản ghi MODS bao gồm ngày tạo, các chỉ số điều khiển.

Mặc dù được sinh ra từ MARC21 và chi tiết hơn tiêu chuẩn Dublin Core, nhưng MODS lại ít quy tắc biên mục hơn MARC21. Tương tự như Dublin Core, không có nhiều trường phụ thuộc và tất cả các trường đều có thể lập lại. MODS chứa đựng nhiều giá trị từ MARC, song nó cũng có những khác điểm khác với MARC21: Không có những khái niệm về "dẫn mục chính" ("main entry") hoặc "dẫn mục phụ" ("added entry"), tất cả tác giả đơn giản chỉ là "tác giả"; và một biểu ghi có thể có nhiều nhan đề mà không có một "nhan đề chính" ("main title"). Khi những biểu ghi MARC21 được biên dịch tới MODS, thì có một biểu ghi trong XML mang dấu ấn của MARC. Biểu ghi MODS có thể được tạo ra từ siêu dữ liệu thư tịch mà không khởi đầu từ mục lục thư viện, ví dụ như trích dẫn của bài báo và nó thường được sử dụng trong cơ sở dữ liệu có sự trộn lẫn của mục lục thư viện và dữ liệu biên mục khác.

Tiếp theo trang 17