

# TÀI LIỆU SỐ VÀ MỘT SỐ CHUẨN MÔ TẢ TÀI LIỆU SỐ

*ThS. Đinh Thúy Quỳnh\**  
*ThS. Hoàng Thúy Phương\*\**

**Tóm tắt:** Ngày nay, sự phát triển của khoa học công nghệ đã mở ra một hướng đi mới, đầy triển vọng cho việc lưu trữ, xử lý thông tin của các cơ quan thông tin thư viện. Việc sử dụng kỹ thuật số để biểu diễn thông tin đã dẫn đến sự xuất hiện của một loại hình tài liệu mới, đó là tài liệu số. Với nhiều ưu điểm nổi trội, nguồn tài liệu này đang góp phần làm thay đổi về chất hoạt động thông tin – thư viện. Tuy nhiên, để xây dựng một thư viện số hay xa hơn là việc chia sẻ, khai thác thông tin giữa các thư viện trong nước cũng như trên thế giới thì bên cạnh những yêu cầu về cơ sở vật chất, hệ thống mạng, nguồn lực thông tin, đội ngũ cán bộ,....các cơ quan thông tin thư viện cũng cần phải áp dụng các chuẩn nghiệp vụ tiên tiến phù hợp với điều kiện thực tế. Bài viết giới thiệu các khái niệm, vai trò và một số chuẩn mô tả cho tài liệu số.

**Từ khóa:** Tài liệu số; Mô tả tài liệu số

## **1. Tổng quan vấn đề nghiên cứu và mục tiêu nghiên cứu**

### **1.1. Tổng quan vấn đề nghiên cứu**

Tài liệu số nói chung và các chuẩn mô tả tài liệu số nói riêng là vấn đề hiện nay đang được rất nhiều nhà thư viện học và thông tin học quan tâm. Nhiều tài liệu, bài báo, tạp chí trong và ngoài nước đã đề cập đến vấn đề này ở các góc độ khác nhau như biên mục, chuẩn hóa, chia sẻ, khai thác, ứng dụng,... Đặc biệt, một số hội thảo xoay quanh chủ đề này cũng đã được tổ chức như Hội thảo “MARC 21” tổ chức tại Thành phố Hồ Chí Minh năm 2000, Hội thảo “Xây dựng khổ mẫu MARC Việt Nam” tổ chức tại Trung tâm Thông tin Khoa học và Công nghệ Quốc gia năm 2001, Hội nghị quốc tế “Thư viện Việt Nam hội nhập và phát triển” tổ chức tại Thành phố Hồ Chí Minh vào ngày 28 – 30/8/2006, Hội nghị Quốc tế “Thư viện số châu Á” được tổ chức tại Hà Nội tháng 12/2007, Hội thảo “Phổ biến kiến thức và áp dụng tiêu chuẩn, chuẩn hoá nghiệp vụ cho hội viên trong hoạt động thư viện ở Việt Nam” tổ chức ngày 27/11/2015 tại Thư viện Quốc gia Việt Nam,...

### **1.2. Mục tiêu nghiên cứu**

Bài viết nhấn mạnh tầm quan trọng của tài liệu số cũng như giới thiệu một số chuẩn mô tả tài liệu số nhằm góp phần giúp các thư viện so sánh, đánh giá và lựa chọn được các chuẩn phù hợp khi xây dựng thư viện số.

## **2. Phương pháp nghiên cứu**

---

\* Khoa Thư viện – Thông tin, Trường Đại học Văn hóa Hà Nội

\*\* Trung tâm Thông tin-Thư viện, Trường Đại học Kinh tế Quốc dân

Bài viết sử dụng phương pháp tổng hợp, phân tích tư liệu; phương pháp quan sát, đối chiếu, so sánh.

### **3. Kết quả nghiên cứu**

#### **3.1. Vai trò của tài liệu số trong hoạt động thông tin – thư viện**

Tài liệu số là gì? Cho đến nay, đã có không ít định nghĩa về tài liệu số được đưa ra. Điểm chung của các định nghĩa này đều cho rằng: tài liệu số là tài liệu mà thông tin chứa đựng trong đó đã được biểu diễn dưới dạng mã nhị phân, tức là mã chỉ gồm hai số 0 và 1. Nói cách khác, những thông tin về tài liệu, một phần hay toàn bộ nội dung của tài liệu đã được chuyển thành các bit thông tin dữ liệu và được lưu trữ, khai thác trên máy vi tính, với sự hỗ trợ của một hay một vài thiết bị chuyên dụng, phần mềm ứng dụng và hệ thống mạng máy tính. Tài liệu được tạo lập trực tiếp từ máy vi tính, được lưu trữ thành các tệp dữ liệu (file), với những định dạng khác nhau như: .doc, .exe, .jpg..., và các tài liệu có nội dung là kết quả của quá trình số hoá các loại hình tài liệu khác: giấy, ảnh, phim... đều là tài liệu số. Như vậy, tài liệu số là vật mang tin mà thông tin trong đó được tạo lập bằng phương pháp dùng tín hiệu số hình thành trong quá trình hoạt động của cơ quan, tổ chức, cá nhân. Với cách hiểu đó, có thể coi nguồn tài liệu số là tập hợp có tổ chức những bộ sưu tập thông tin số và thông tin trong tài liệu đã được số hoá.

Tài liệu số có những đặc trưng cơ bản như mật độ thông tin lớn; cập nhật nhanh chóng, có khả năng truy cập theo nhiều dấu hiệu khác nhau, bởi nhiều người dùng ở cùng một thời điểm mà không bị giới hạn về thời gian hay vị trí địa lý; có thể lưu trữ theo nhiều định dạng khác nhau,... Chính những đặc trưng này đã làm cho nguồn tài liệu số ngày càng giữ vai trò quan trọng trong hoạt động giao lưu thông tin nói chung và hoạt động thông tin – thư viện nói riêng. Chúng ta có thể kể đến các vai trò đó là:

- *Tài liệu số là “hạt nhân” của Thư viện số*

Thư viện số là một kho thông tin số hoá, được cấu trúc để dễ dàng truy cập thông qua các mạng máy tính hay các mạng viễn thông.

Số hoá tài liệu và xây dựng bộ sưu tập số là lựa chọn tối ưu để bảo tồn lâu dài các tài liệu gốc quý hiếm, ngăn chặn những rủi ro, huỷ hoại do thời gian, thiên tai... Đối với hoạt động đào tạo và học tập trong trường đại học, bộ sưu tập số cũng góp phần phục vụ hiệu quả cho công cuộc đổi mới và nâng cao chất lượng đào tạo, đặc biệt là đào tạo theo tín chỉ, đào tạo trực tuyến... Ngoài ra, nhờ đó, việc “học tập suốt đời” (long-life study) được khuyến khích vì người học có thể tham khảo tài liệu ở bất kỳ đâu, bất kỳ thời gian nào thuận tiện với họ.

Phần cốt lõi của Thư viện số là kho tài liệu số hay tập hợp của các bộ sưu tập số. Vì vậy, việc phát triển kho tài liệu số là công việc quan trọng hàng đầu trong xây dựng Thư viện số.

- *Tài liệu số là cơ sở của sự hình thành và phát triển ngành xuất bản điện tử*

Tài liệu số ra đời đã góp phần làm thay đổi diện mạo ngành xuất bản với hàng loạt sản phẩm mới được tung ra thị trường với tên gọi loại hình là Ấn phẩm điện tử hay Xuất bản phẩm điện tử. Ý tưởng đầu tiên về xuất bản điện tử được ông Vanner Bush đưa ra tại một hội thảo khoa học ở Viện Công nghệ Massachusertte (Mỹ) vào năm 1945. Sau đó ít lâu, đã có một sự bùng nổ hàng loạt các sản phẩm kỹ thuật số và các lợi thế về kỹ thuật không thể phủ nhận. Năm 1991, cuốn sách điện tử đầu tiên chào đời tại Mỹ. Từ năm 1998, sách điện tử được xuất bản ở dạng đĩa và có thể được tải xuống máy tính cá nhân từ Internet.

Các loại hình xuất bản phẩm điện tử có trên thị trường hiện nay là sách điện tử (E-Book) và các báo, tạp chí điện tử (E-Journal).

- *Tài liệu số là cơ sở của sự hình thành và phát triển ngành công nghiệp nội dung số*

Sự phát triển có tính chất bùng nổ của các nguồn tài liệu số trong những năm gần đây đã dẫn đến hình thành khái niệm “*nội dung số*” và kéo theo đó là sự ra đời của ngành công nghiệp nội dung số.

Các sản phẩm chứa đựng nội dung số, khi tham gia thị trường, sẽ làm hình thành nên ngành công nghiệp nội dung số (CNNDS). Theo quan điểm của Bộ Bru chính- Viễn thông Việt Nam: CNNDS là ngành thiết kế, sản xuất, xuất bản, lưu trữ, phân phối, phát hành các sản phẩm nội dung số và dịch vụ liên quan. Tập hợp các hoạt động liên quan đến số hoá thông tin ở quy mô lớn, có sản phẩm hoàn chỉnh và có khả năng đem lại lợi nhuận thì được gọi là ngành CNNDS. Nguyên vật liệu đầu vào của ngành CNNDS là thông tin và sản phẩm đầu ra cũng là thông tin. Nhưng thông tin đầu ra là CSDL, là thông tin có cấu trúc, có nội dung cụ thể và được cung cấp cho những đối tượng cụ thể nhằm phục vụ cho một hay một số hoạt động kinh tế - xã hội nhất định.

### **3.2. Một số chuẩn mô tả tài liệu số**

#### ***MARC (MACHine READable Cataloging - Biên mục đọc máy)***

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. Nó 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. 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 con 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 trường con.

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 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 21 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 (gồm hai nhóm là trường kiểm soát và trường dữ liệu): là những trường dữ liệu có trong biểu ghi và chứa các dữ liệu mô tả. Các trường dữ liệu là trường có độ dài biến động (Variable Fields).

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.

Tính ưu việt và tiêu chuẩn của MARC:

- Khổ mẫu tích hợp, độ dài trường dữ liệu có thể thay đổi, các trường lặp và tổ chức các điểm truy cập.
- Cấu trúc khổ mẫu MARC là cấu trúc biểu ghi, đó là một triển khai ứng dụng tiêu chuẩn quốc tế ISO 2709 - Khổ mẫu trao đổi thông tin (Format for Information Exchange) và tiêu chuẩn tương đương của Hoa Kỳ ANSI/NISO Z39.2 - Trao đổi thông tin thư mục (Bibliographic Information Interchange).

Hiện nay, MARC đang phát huy ảnh hưởng trong mọi hoạt động thư viện bao gồm: chọn lọc, đặt mua, biên mục, tìm tin, biên soạn và xuất bản thư mục,... Khổ mẫu này đang phục vụ đắc lực cho việc nối mạng và truy cập thông tin trực tuyến.

### **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 MARC 21 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.

Bộ yếu tố siêu dữ liệu Dublin Core được thiết kế chủ yếu cho mục đích mô tả bao gồm 15 yếu tố sau:

STT	Các yếu tố DC	Mô tả
1	Nhan đề (Title)	Nhan đề của tài liệu
2	Tác giả (Creator)	Tác giả của tài liệu, bao gồm cả tác giả cá nhân và tác giả tập thể
3	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)
4	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..
5	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..
6	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..
7	Ngày tháng (Date)	Ngày, tháng ban hành tài liệu. Có thể dùng chuẩn ISO 8601 ( <a href="http://www.w3.org/TR/NOTE-datetime">http://www.w3.org/TR/NOTE-datetime</a> )
8	Loại (kiểu) (Type)	Mô tả bản chất của tài liệu. Dùng các thuật ngữ mô tả phạm trù kiểu: trang chủ, bài báo, báo cáo, từ điển...
9	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....)
10	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).
11	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

- |    |                          |   |
|----|--------------------------|---|
| 12 | 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                         |
| 13 | Liên kết (Relation)      | Mô tả các thông tin liên quan đến tài liệu khác. có thể dùng đường dẫn (URL), URN, ISBN, ISSN.  |
| 14 | 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, tọa độ.. |
| 15 | Bản quyền (Right)        | Các thông tin liên quan đến bản quyền của tài liệu  |

Mục đích của Dublin Core là cung cấp một bộ phân 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 có được sự mô tả thư tịch riêng của chúng và chúng sẽ có những phân tử dữ liệu được mã hóa 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 đã 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 đã cung cấp một bộ phân 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 hóa khác nhau do đó làm tăng tính mở giữa các cộng đồng khi muốn trao đổi và chia sẻ tài nguyên cho nhau.

Dublin Core có những ưu điểm sau:

- Đơn giản trong tạo lập và bảo trì: được thiết kế nhằm phục vụ những người không chuyên; dễ sử dụng và rẻ nhưng mang lại hiệu quả lớn.
- Ngữ nghĩa thông dụng: khắc phục những khó khăn trong việc hiển thị các thuật ngữ. VD: yếu tố <tác giả> (Creator) được gán cho người tạo lập, nhà soạn nhạc, đạo diễn, trong vai trò là tác giả chính.
- Phạm vi quốc tế: tháng 11 - 1999, đã có phiên bản của hơn 20 thứ tiếng: Phần Lan, Na Uy, Thái Lan, Nhật, Pháp, Đức, Hy Lạp, Indonesia, Tây Ban Nha. Tổ chức WWW phát triển Chuẩn Dublin core trên nền tảng kết hợp đa ngôn ngữ, phục vụ cho môi trường tài nguyên thông tin điện tử mang tính chất đa văn hoá và đa ngôn ngữ.
- Khả năng mở rộng: với cơ chế mở, Chuẩn Dublin core có thể được mở rộng bởi các chuyên gia bằng việc gắn kết thêm các yếu tố mở rộng. Khả năng này còn được thực hiện một cách đơn giản thông qua việc kết nối nhiều CSDL khác nhau thông qua mạng Internet

***METS (Metadata Encoding and Transmission Standard - Tiêu chuẩn về mã hóa và truyền siêu dữ liệu)***

Trong lĩnh vực thư viện số, sự đa dạng về siêu dữ liệu lớn hơn rất nhiều. Các chương trình phát triển thư viện số luôn cần có một số những tiêu chuẩn mở rộng, đặc biệt cho việc mô tả siêu dữ liệu sao cho chính xác và gần gũi hơn. Một chuẩn chung được coi như khung *framework* cho các siêu dữ liệu đó chính là METS. METS hoạt động như một nền tảng mà trong đó các tiêu chuẩn trước đó cũng như các tiêu chuẩn trong tương lai có thể được tích hợp một cách logic và có cấu trúc.

METS là một tiêu chuẩn mới được thiết kế để mã hóa 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. Cho đến gần đây, chưa có phương pháp chuẩn hóa nào cho việc mã hóa 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 họ, 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)

Chuẩn này là sự tổng hợp tất cả các loại hình siêu dữ liệu: 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. Các tài liệu METS có các phần và các yếu tố liên quan hầu hết đến các đối tượng số, ngoài ra nó có khả năng mô tả các đối tượng xác định bao gồm việc tham chiếu đến metadata mở rộng theo các định dạng được chuẩn hóa.

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

STT	Thành phần	Mô tả
1	Đầu mục METS (METS Header)	Bao gồm metadata về bản thân tài liệu của METS file như: tác giả, người tạo, ngày tháng, sản phẩm, trạng thái...;
2	Siêu dữ liệu mô tả (Descriptive Metadata)	Có một hoặc nhiều trường hợp của metadata mô tả về đối tượng TVS, có thể xem xét metadata ngoài tài liệu METS hoặc bên trong hoặc cả hai
3	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;
4	File tóm tắt (File Inventory)	Danh sách tất cả các file gồm nội dung thực sự cho đối tượng TVS;
5	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 các liên kết các thành phần của cấu trúc tới các file nội

dung;

- |   |                                      |  |
|---|--------------------------------------|--|
| 6 | 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 structural map   |
| 7 | 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 có phải có 2 loại chuẩn hóa: chuẩn hóa siêu dữ liệu của chính vật chứa (giống như định dạng MARC trong lĩnh vực thư viện truyền thống), và việc chuẩn hóa 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.

Những ưu điểm của METS:

- Nó có khả năng lưu trữ mạnh mẽ, con người có thể đọc được, và luôn có khả năng trao đổi với các ứng dụng tương thích XML.
- METS được thiết kế một cách rõ ràng rành mạch cho rất nhiều loại siêu dữ liệu mà có thể áp dụng được cho các đối tượng của thư viện số. METS có cấu trúc chặt chẽ có thể dễ dàng được tự động tạo ra từ các ứng dụng khác (như là các CSDL quan hệ).
- METS có khả năng cung cấp cho người sử dụng 2 sự lựa chọn: nhúng trực tiếp siêu dữ liệu vào cấu trúc hoặc tham chiếu tới các siêu dữ liệu được lưu trữ từ bên ngoài. Điều này giúp cho METS có sự linh hoạt lớn và khá dễ dàng trong việc chuyển đổi các siêu dữ liệu cũ.
- METS cũng được thiết kế để có khả năng sẵn sàng mở rộng vì thế các yêu cầu mới được cộng đồng thư viện số đưa ra có thể được kết hợp một cách nhanh chóng theo một quy trình xử lý đơn giản. Vì thế METS có khả năng phát triển để có thể trợ giúp cho các thay đổi trong tương lai của các hoạt động trong thư viện số.

Tuy nhiên chính sự linh hoạt của METS cũng tạo một số khó khăn cho việc ứng dụng METS một cách rộng rãi: Nếu nội dung của siêu dữ liệu chưa được chuẩn hóa thì vật chứa các siêu dữ liệu đó cũng không có nhiều giá trị. Trong thư viện truyền thống, MARC được hỗ trợ bởi quy tắc biên mục Anh – Mỹ, quy tắc cung cấp những chỉ dẫn cho nội dung của biểu ghi thư mục. Và METS không phản ánh nội dung của siêu dữ liệu được lưu trữ trong METS, điều này khác với MARC làm đối với siêu dữ liệu của MARC. Ban biên tập METS đang cố gắng giải quyết vấn đề này bằng cách đưa ra các giản đồ mở rộng

sử dụng cùng với METS. Đó là các lược đồ XML bên ngoài đơn giản, được mã hóa bởi các siêu dữ liệu phổ biến như Dublin Core, và có khả năng liên kết với gián đồ METS bằng các kỹ thuật sử dụng các XML phổ biến gọi là Phần mở rộng. Tuy nhiên, cũng giống như METS, trong nhiều trường hợp các gián đồ này cũng chỉ là các vật chứa siêu dữ liệu.

Ở một số khía cạnh nào đó thì sự linh hoạt của METS cũng là điểm yếu nhất của METS: Nó chỉ là một phần trong giải pháp giải quyết vấn đề về chuẩn hóa siêu dữ liệu trong thư viện số.

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

Trong một môi trường nối kết 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 là mọi hệ thống có thể cần và 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 thẻ 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 thẻ 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; nó cũng cho phép lưu lại các đối tượng liên quan để có thể tương tác theo cách đó. Ngoài ra, nó còn có khả năng dễ dàng trong việc 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, nó 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 MARCXML.

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. Nó 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 metadata 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.

Cấu trúc của MODS: gồm 19 thành tố chính sau:

STT	Thành phần chính	Mô tả
1	Thông tin nhan đề (TitleInfo)	Là một yếu tố bắt buộc, yếu tố này gồm 5 yếu tố con, một trong số đó là title, đây là thành phần bắt buộc. title được sử dụng để ghi lại tiêu đề chính cho item 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.
2	Tên (Name)	Yếu tố này tương đương với trường Tác giả ( <i>Creator</i> ) và Tác giả phụ ( <i>Contributor</i> ) trong Dublin Core, name 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).
3	Thông tin gốc (OriginInfo)	Là một yếu tố gộp khác, mang thông tin tổng hợp có nguồn gốc hoặc phổ biến của biểu ghi. 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.
4	Mô tả vật lý (PhysicalDescription)	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 internetMediaType, là thành phần ghi lại định dạng dữ liệu được mô tả (thường có định dạng theo chuẩn MIME như "text/html"), reformattingQuality (về độ phân giải và số màu) khi tài liệu được scan, digitalOrigin ghi lại xem đối tượng nào được số hóa 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 extent ghi lại số trang, và yếu tố note ghi lại các thông tin không cấu trúc về các đặc tính vật lý của đối tượng.
5	Chủ đề (Subject)	Là một yếu tố gộp được sử dụng để mô tả đề mục của item theo các quy tắc phân loại nào đó. Yếu tố con hierarchicalGeographic có thể dùng để định nghĩa sự phân cấp của các thuật ngữ 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 cartographics, cho phép lưu lại chi tiết theo tọa độ địa lý và theo phạm vi trong bản đồ.
6	Tài liệu liên quan (RelatedItem)	Đây là một yếu tố rất hữu ích trong ngữ cảnh của bộ sưu tập các tài liệu có một số mối liên hệ với nhau (như bộ sưu tập các bài báo và tuần san được số hóa), relatedItem cho phép các item 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 (trong Dublin Core, nhưng có khả năng mềm dẻo hơn trong cách sử dụng).
7	Mở rộng (Extension)	Mặc dù MODS cung cấp tập các yếu tố mở rộng hơn Dublin Core, nhưng nó vẫn không thể làm thỏa mãn tất cả các yêu cầu metadata cho đối tượng. Trong trường hợp đó, nó cung cấp một cơ chế dễ dàng mở rộng tập các yếu tố bằng cách cho phép ghi lại metadata trong các lược đồ có khả năng thay đổi được gắn vào trong tài liệu MODS record.
8	Thuộc tính của nguồn (TypeOfResource)	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.
9	Thẻ loại (Genre)	Là một quy định đặc biệt hơn typeOfResource, yếu tố này cho phép quy định phân loại chi tiết để mô tả cho item.
10	Ngôn ngữ (Language)	Ghi lại ngôn ngữ mô tả của item
11	Tóm tắt (Abstract)	Mô tả nội dung tóm tắt của item hoặc một liên kết tới một định dạng tương tự như vậy.
12	Mục lục (TableOfContents)	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.
13	Người sử dụng mục tiêu (TargetAudience)	Quy định các đối tượng người sử dụng (ví dụ như vị thành niên, hoặc thanh niên) cho tài liệu.
14	Ghi chú (Note)	Là yếu tố ghi lại các thông tin cần lưu ý.
15	Phân loại (Classification)	Số phân loại tài nguyên theo một lược đồ đã được thông qua ví như Tiêu đề trong thư viện quốc hội Mỹ hoặc phân loại số thập phân Dewey.

16	Định danh (Identifier)	Số hoặc mã duy nhất phù hợp với lược đồ đã được thông qua như số ISBN hoặc ISS
17	Vị trí (Location)	Ghi lại vị trí vật lý của tài liệu, bao gồm kho lưu trữ và chỉ số giá của item
18	Hạn chế truy cập (AccessRestriction)	Thông tin về việc truy cập item bị hạn chế thế nào, bao gồm thông tin về bản quyền
19	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 nhiều tiêu chuẩn Dublin Core, những 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 lặp lại. MODS chứa đựng nhiều giá trị từ MARC, xong nó cũng có những nhiều 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 người ta 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.

#### 4. Kết luận

Như vậy, chúng ta có thể thấy, hiện nay trên thế giới có khá nhiều chuẩn mô tả dữ liệu khác nhau. Chẳng hạn như chuẩn biên mục tự động trên máy tính MARC và Dublin Core quy định về cơ cấu và cấu trúc dữ liệu cho biểu ghi trong các cơ sở dữ liệu của một hệ thống thông tin cụ thể, trong đó MARC quy định 200 trường mô tả dữ liệu còn Dublin Core là một chuẩn ra đời sau và chỉ quy định 15 trường rút gọn, đơn giản, phù hợp với các tài liệu trên môi trường Internet hiện nay, còn khi các tài nguyên số phát triển và mở rộng về dữ liệu đa phương tiện thì các chuẩn mới là METS và MODS cũng được phát triển và ứng dụng...Do vậy, tùy từng nhu cầu sử dụng và quy mô của thư viện mà chúng ta sẽ đưa ra những quyết định trong việc lựa chọn các chuẩn mô tả sao cho phù hợp.

#### TÀI LIỆU THAM KHẢO

##### Tiếng Việt

1. Nguyễn Tiên Đức (2005), “Xây dựng thư viện điện tử và vấn đề số hoá tài liệu ở Việt Nam”, *Thông tin và Tư liệu*, (2).
2. Nguyễn Minh Hiệp (2004), “Thế giới Thư viện số”, *Thư viện – Công nghệ thông tin*, (4), tr. 2 – 13.
3. *Khổ mẫu MARC 21 cho dữ liệu thư mục* (2005), Trung tâm Thông tin Khoa học Công nghệ Quốc gia, Hà Nội
4. Cao Minh Kiêm (2000), “Thư viện số: Định nghĩa và vấn đề”, *Thông tin và Tư liệu*, (3), tr. 5 – 11.

##### Tiếng Anh

5. Linda Cantara (2005), *METS: The Metadata Encoding and Transmission Standard*, truy cập ngày 25/11/2016, tại địa chỉ trang web: <http://www.columbia.edu/cu/libraries/inside/units/bibcontrol/osmc/cantara.pdf>
6. Stuart L. Weibel (2000), "The Dublin Core Metadata Initiative Mission, Current Activities, and Future Directions", truy cập ngày 25/11/2016, tại địa chỉ trang web: <http://www.dlib.org/dlib/december00/weibel/12weibel.html>
7. H. Witten, D. Bainbridge (2002), *How to Build a Digital Library*, Elsevier Science Inc., NY.
8. <http://www.loc.gov/standards>
9. <http://dublincore.org/>