

TỪ MARC ĐẾN MARC-XML VÀ NON-MARC METADATA

ThS. NGUYỄN MINH HIỆP

Thư viện ĐH Khoa học Tự Nhiên, ĐHQG TP. Hồ Chí Minh

*Mạng toàn cầu Internet ra đời vào thập niên 1960; cuộc bùng nổ thứ nhất về kết nối xảy ra trong thập niên 1980; và sau khi Tim Berners Lee phát minh Web vào năm 1991, cuộc bùng nổ thứ hai về sử dụng World Wide Web đã đưa ngành quản lý thông tin đạt đến đỉnh cao. Tại thời điểm này, 1995, công nghệ Web (Công nghệ IP-based - Sử dụng HTTP trong việc chuyển thông tin và HTML/XML trong việc đóng gói thông tin) hoàn toàn chi phối ngành thông tin thư viện, nhất là sau khi thư viện số ra đời và trở nên thịnh hành khắp nơi trên thế giới. Vai trò vinh quang của MARC trong giai đoạn đầu tin học hóa phải nhường chỗ cho Web. Lần đầu tiên trong lịch sử biên mục, người ta dùng ngôn ngữ XML. Kể từ đó **Web chính là công nghệ của ngành thông tin thư viện trong hiện tại và tương lai.***

Sự ra đời của MARC

Chuẩn hóa thư viện gắn liền với lịch sử biên mục. Tiêu chuẩn đã từng nổi tiếng nhất trong lịch sử biên mục chính là phiếu mục lục 3x5 inch. Phiếu mục lục được viết bằng tay, đánh máy, và in bằng vi tính; mô tả phiếu có thể khác nhau nhưng kích cỡ phiếu đục lỗ 3x5 thì đồng nhất với nhau trên toàn thế giới. Hệ thống mục lục phiếu tồn tại trong một thời gian dài trong lịch sử phát triển thư viện. Những tiêu chuẩn về mô tả như ISBD, AACR2 thúc đẩy tiến trình chuẩn hóa trong công tác biên mục. Hơn nữa khi các thư viện bắt đầu nói chuyện với nhau thì

việc hợp tác biên mục được đặt ra và nảy sinh ra việc sao chép phiếu mục lục để trao đổi với nhau hơn là phải mỗi thư viện làm công việc biên mục giống nhau cho cùng một tài liệu. Như thế các thư viện cần phải chuẩn hóa để đáp ứng nhu cầu trao đổi và chia sẻ biểu ghi mục lục (phiếu mục lục). Việc chia sẻ biểu ghi đòi hỏi ít nhất hai mức độ chuẩn hóa: nội dung và hình thức (hay là ngữ nghĩa và cú pháp). Thư viện Quốc hội Hoa Kỳ dần dần đảm nhận việc cung cấp bộ phiếu mục lục cho các thư viện khắp nơi thông qua các nhà phân phối sách.

Máy tính xuất hiện và được áp dụng ngay vào công tác biên mục. Mục lục máy tính đã thu hút cán bộ biên mục vì năng suất chỉ mục tăng vọt. Cũng như những bộ phiếu in, Thư viện Quốc hội đã tải xuống (download) những biểu ghi lên băng từ để phân phối – nhưng chỉ cho những thư viện có khả năng đọc và tải lên (upload) vào hệ thống thư viện tự động hóa của mình. Do đó một tiêu chuẩn hình thức mới được yêu cầu – và MARC (MACHine Readable Cataloging) ra đời để đáp ứng yêu cầu đó.

Vài nét lịch sử của MARC

Bản chất của MARC là một dạng thức trao đổi biểu ghi. Sự phát triển của MARC liên quan mật thiết đến tiến trình phát triển tự động hóa thư viện và việc cải biên mục lục sao chép.

Thư viện Quốc hội Hoa Kỳ phát triển MARC I trong những dự án đầu tiên trong khoảng 1965-1968, rồi hình thành MARC II, 1968 và trở thành USMARC. Tuy nhiên với USMARC chưa phải đã là kết thúc. Trong khi ở Hoa Kỳ, Thư viện Quốc hội tiếp tục chiếm lĩnh việc phân phối biểu ghi mục lục, thì những nơi khác nhiều cơ sở cũng tự mình hình thành những nhà cung cấp. MARC có một tiêu chuẩn hình thức phức tạp hơn phiếu đục lỗ 3x5 inch nhiều – nó phụ thuộc nhiều hơn vào nội dung biểu ghi và nội dung này khác với nội dung được cung cấp từ Thư viện Quốc gia Anh và những thư viện khác. Do đó, Thư viện Quốc gia Anh theo hướng dẫn của Thư viện Quốc hội Hoa Kỳ nhưng không chấp nhận MARC II cho

dịch vụ phân phối MARC của chính họ; thay vào đó, UKMARC được phát minh và được sử dụng vào việc sao chép biểu ghi đối với những thư viện phụ thuộc vào Thư viện Quốc gia Anh. Những biến thể MARC khác được hình thành với lý do tương tự: AUSMARC ở Úc, MAB ở Đức, CanMARC ở Canada, INTERMARC ở Pháp. Những hệ thống thư viện thương mại cung cấp những biến thể MARC theo yêu cầu khác nhau của những quốc gia trên thế giới.

Mặc dù những băng từ của Thư viện Quốc hội Hoa Kỳ có thể được gửi đến thư viện của những quốc gia khác, trong thập niên 1970 và 1980 những quốc gia ngoài nước Mỹ phụ thuộc vào thư viện quốc gia và những công-xoóc-xi-um địa phương vẫn thích dùng những tiêu chuẩn thư tịch địa phương hơn. Do đó nhiều MARC quốc gia tiếp tục ra đời, chẳng hạn như NORMARC 1971 (Na Uy), IBERMARC 1976 (Tây Ban Nha), DANMARC 1976 (Đan Mạch), LIBRISMARC 1976 (Thụy Điển), FINMARC, 1977 (Phần Lan), MALMARC 1978 (Malaysia), PHILMARC 1979 (Philippin), KORMARC 1981 (Hàn Quốc), Japan/MARC 1981 (Nhật Bản), Chinese MARC 1982 (Đài Loan), SAMARC 1982 (Nam Phi), New Zealand MARC 1982 (Tân Tây Lan), SBN/ANAMARC 1983 (Italia), SINGMARC 1982 (Singapore), Indian MARC 1985 (Ấn Độ), PUL MARC 1985 (Trung Quốc), INDOMARC 1990 (Indonesia), vv... Tuy nhiên, khi công nghệ xuất bản trở nên mở rộng trên phạm

vi quốc tế, nhiều thư viện ủng hộ việc quốc tế hóa chuẩn MARC. Do đó UNIMARC do IFLA thành lập vào giữa thập niên 1970 nhằm đáp ứng nhu cầu đồng bộ hóa. Tuy nhiên UNIMARC ra đời quá trễ: USMARC đã trở nên quá vững chắc ở Mỹ và những cơ sở dữ liệu khổng lồ ở Mỹ có ảnh hưởng lớn đến nhiều nước trên thế giới đều dùng USMARC. Nếu những thư viện Hoa kỳ không chấp nhận UNIMARC, thì các thư viện Úc, Tân Tây Lan, Canada, và Anh sẽ phải chấp nhận USMARC.

Từ đầu thập niên 1990, nhiều nước trên thế giới lần lượt từ bỏ MARC quốc gia để chuyển sang dùng USMARC. Để mang tính quốc tế hơn, USMARC đã phối hợp với CanMARC của Canada để tạo nên MARC 21. Anh từ bỏ UKMARC chuyển sang MARC 21. Ít lâu sau Úc từ bỏ AUSMARC chuyển sang MARC 21. Vì lợi ích kinh tế và nhu cầu trao đổi thông tin với tất cả những cơ sở dữ liệu lớn trên thế giới, tất cả những quốc gia nói tiếng Anh và ngay cả nhiều quốc gia không nói tiếng Anh đã chuyển từ MARC quốc gia sang MARC 21. Trước bối cảnh đó và trong cố gắng hội nhập với hệ thống thông tin - thư viện thế giới hiện đại, Nga cho ra đời RUSMARC, 1998 khá muộn màng và lẻ loi. Hơn thế nữa, tại thời điểm đó, giá trị của MARC đã thay đổi lớn. Nhu cầu trao đổi thông tin không dừng lại ở chỗ chỉ trao đổi biểu ghi thư tịch mà còn dưới dạng toàn văn, âm thanh, hình ảnh, hình ảnh động, vv... MARC không thích hợp để làm điều đó, một công nghệ hoàn toàn khác hẳn đã thay thế, đó là công nghệ IP-

based với ngôn ngữ XML. Hội nghị tại thành phố Dublin, bang Ohio, Hoa kỳ vào năm 1995 cho ra đời chuẩn Dublin Core đã chính thức đặt nền móng cho việc sang trang lịch sử biên mục học – biên mục trên nền Web và ấn định việc trao đổi dữ liệu bằng metadata. Chuẩn Z39.50 không thích hợp và chức năng trao đổi biểu ghi thư tịch của MARC không còn đóng vai trò chủ đạo.

Sau hội nghị DCMI (Dublin Core Metadata Initiative) đầu tiên tại Dublin, đã có nhiều hội nghị được tổ chức trên bốn châu lục để hoàn thiện chuẩn biên mục Dublin Core. Mười lần hội nghị được tổ chức tại Anh, Úc, Phần Lan, Đức, Canada, Nhật, Italia, và Hoa Kỳ. Tháng 9/2003 sẽ tổ chức hội nghị tại Seattle, Hoa Kỳ và năm 2004 sẽ tổ chức tại một quốc gia Châu Á. Điều này đã giải quyết dứt điểm việc chuẩn hóa trao đổi biểu ghi thư tịch nói riêng và thông tin nói chung. Đồng thời vai trò vinh quang của MARC cũng âm thầm kết thúc. Chưa bao giờ ngành thông tin thư viện trên thế giới có một thuận lợi rất lớn để tiến đến chuẩn hóa việc trao đổi thông tin trên phạm vi toàn cầu như hiện nay. Cho dù đang dùng MARC 21, UNIMARC, MARC quốc gia hay chưa hề dùng MARC điều đó không quan trọng đối với công nghệ mới.

Trước bối cảnh đó thật ngạc nhiên khi tại nước ta vào năm 2002 lại một sản phẩm MARC quốc gia ra đời đó là VNIMARC hay còn được gọi là MARC 21 Việt Nam !

MARC-XML

Chuẩn phiếu mục lục 3x5 inch và chuẩn MARC là hai cây cổ thụ đã mang đến những vinh quang rất lớn trong lịch sử biên mục học thế giới. Nhưng theo quy luật, cái mới phải thay cái cũ. Sự suy tàn của cả hai đánh dấu hai mốc phát triển quan trọng trong lịch sử biên mục:

- một, khi máy tính ra đời – hệ thống mục lục máy tính và trực tuyến thay cho hệ thống mục lục phiếu;
- hai, khi bùng nổ việc sử dụng thông tin điện tử hay chính xác hơn khi thư viện số ra đời – biểu ghi mục lục được số hóa, rồi chính tài liệu được số hóa đòi hỏi truy cập qua hệ thống mục lục.

Sự suy tàn của những cây cổ thụ phải theo cách của "đại thụ". Khi chuyển từ hệ thống mục phiếu sang hệ thống mục lục máy tính, người ta không muốn phải dứt bỏ ngay những tủ phiếu mà ta đã dày công tạo dựng từ năm này sang năm khác mà viện đủ lý do để giữ lại, mặc dù nhiều lúc tỏ ra là thừa. Ngoài ra nhiều thư viện vẫn phải giữ lại những tủ phiếu đó một khi chưa hồi cố hết số sách trong kho chuyên sang tin học hóa, chẳng hạn như tại Thư viện đại học lớn nhất nước Mỹ là Harvard và Thư viện đại học Cambridge của nước Anh vẫn còn sử dụng tủ phiếu để tra cứu những tài liệu cũ. Ngay tại Thư viện Quốc hội Hoa Kỳ cũng có một tòa đại sảnh để chứa hàng trăm tủ phiếu mục lục để ghi nhớ thời vinh quang của nó. Ngày nay số phận MARC cũng tương tự như thế. Ngoài

trừ cách giải quyết khá quyết liệt của ông khổng lồ thứ hai của nước Mỹ là Thư viện Quốc gia Y học đã từ bỏ MARC và chuyển hẳn hơn 10 triệu biểu ghi sang XML từ năm 2001, những thư viện lớn khác kể cả Thư viện Quốc hội Hoa Kỳ theo con đường nước đôi, phần lớn vì lý do kinh phí vẫn còn sử dụng MARC cho những biểu ghi thư tịch kèm theo metadata. Điều này áp dụng cho tất cả các MARC kể cả MARC 21, do đó có tên là MARC-XML hay MARC 21-XML. Những người theo cách này cho rằng *"Chúng ta chưa có nhu cầu từ bỏ truyền thống biên mục phong phú của chúng ta trong khi việc chia sẻ metadata vẫn còn trong trứng nước, như là một phần mềm mà thôi"*.

Một cách khách quan, chúng ta có thể hình dung MARC-XML như hai cái cổng của một ngôi nhà. Cổng MARC kèm theo giao thức Z39.50 là cổng được sử dụng từ trước đến giờ để trao đổi biểu ghi thư tịch – hay nói một cách khác là thông tin dạng thư tịch. Ngày nay yêu cầu trao đổi thông tin không dừng lại ở dạng thư tịch mà cao cấp hơn (hình ảnh, hình ảnh động, âm thanh, điện tử, vv...), cổng MARC không đảm đương được dạng cao cấp đó, người ta phải xây một chiếc cổng khác hiện đại hơn, đó là cổng XML có thể trao đổi được tất cả các dạng thông tin. Cổng XML trở thành cổng chính và dần dần được chuẩn hóa để trở thành tiêu chuẩn cho tất cả các ngôi nhà. Chiếc cổng MARC trở thành cổng phụ, cũng vẫn còn được dùng để tiếp nối việc trao đổi bên ngoài chỉ với thông tin dạng thư tịch –

Giống như ngày xưa có hệ thống tra cứu máy tính rồi nhưng vẫn còn giữ lại tủ phiếu vì có người vẫn chưa quen tra cứu máy tính thậm chí vẫn còn muốn tra cứu trên tủ phiếu.

Chiếc công cụ MARC xem ra là một vấn đề cho những ai đã dày công xây dựng nó. Giá trị mà MARC đã mang lại trong thời gian qua quả thật lớn đã là một vấn đề tâm lý đè nặng khi phải quyết định dứt bỏ nó, ngoài ra chính Thư viện Quốc hội Hoa Kỳ cũng đang còn cu rục mang khiến những người đang dùng MARC càng lưỡng lự.

Thế còn những thư viện chưa hề dùng MARC như chúng ta thì sao? – Một câu trả lời hết sức bình thường rằng chúng ta chỉ việc xây dựng chiếc công cụ chuẩn XML và phát triển nó theo nhịp chung với tất cả mọi người. Đó là hướng non-MARC Metadata.

Non-MARC Metadata

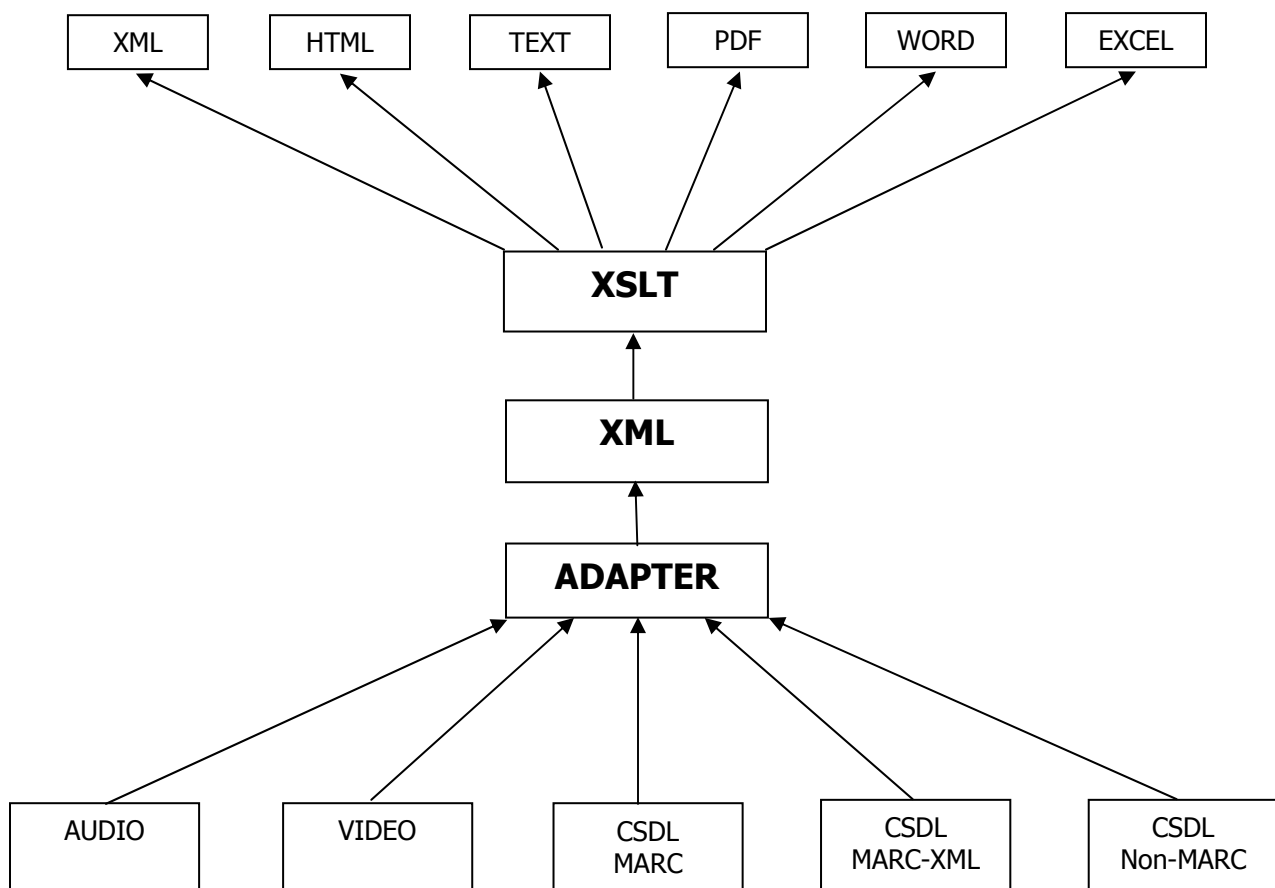
MARC được phát triển bởi thư viện trước hết để phục vụ cho biên mục thư viện. Mục lục trực tuyến đã có một ảnh hưởng rất lớn đối với thư viện hơn 30 năm qua trong việc chia sẻ biểu ghi mục lục ngay cả khi biểu ghi được số hóa. MARC đã đóng vai trò quan trọng trong việc tổ chức hệ thống biên mục đồng nhất nhằm phục vụ việc bổ sung và phát triển tư liệu. Ngày nay tư liệu số hóa không đòi hỏi phải theo phương thức phân phối truyền thống trước đây – người cán bộ biên mục không thể biên mục tài liệu điện tử theo cách của sách, bản đồ, băng video, vv...

Vai trò quan trọng trong việc cung cấp biểu ghi mục lục của Thư viện Quốc hội Hoa Kỳ không còn nữa. Tài liệu để biên mục không phải chỉ có trong thư viện mà có thể đến từ viện bảo tàng, cơ sở lưu trữ, trường đại học, viện nghiên cứu, cơ quan công quyền, vv... Tất cả được đặt trên Internet, được sử dụng rộng rãi, không phải chỉ dành riêng cho thư viện.

Những metadata không thuộc thư viện ngày càng gia tăng về khối lượng hiện hữu và tiềm năng không đòi hỏi phải dựa vào MARC. Bản thân những cơ sở cung cấp tài nguyên thông tin trên mạng Internet có thể tự xuất bản điện tử và cung cấp metadata với dạng thức đơn giản hơn MARC nhiều. Những dạng thức không thuộc thư viện này thông dụng hơn MARC là vì tự chúng đã là máy đọc được rồi.

Những loại hình thư viện khác nhau đòi hỏi những dạng thức khác nhau đối với những loại metadata khác nhau. Họ không cần phải tiếp nhận metadata từ Thư viện Quốc hội Hoa Kỳ, do đó không cần thiết phải chấp nhận MARC; thật ra có một số thư viện chẳng biết MARC là gì.

Non-MARC Metadata chính là chọn lựa của tất cả chúng ta. Sơ đồ dưới đây cho chúng ta thấy rõ cơ chế hoạt động trong việc trao đổi thông tin mà MARC chẳng đóng một vai trò gì trong hệ thống này. Công tác biên mục được thực hiện bởi một công cụ "chuyển đổi nội dung - content transformer" được gọi là adapter. Ảnh hưởng của MARC được dừng lại sau những adapter.



TÀI LIỆU THAM KHẢO

1. Chandler, Adam. Distributing non-MARC metadata: the CUGIR metadata sharing project. *Library Collections, Acquisitions, & Technical Services* 26(2002) 207-217
2. Chen, Hsueh-hua. Metadata : Resource Description for Digital Libraries. *ASIA Library News* 2002 September; Vol. 5 No. 3
3. Cover, Robin. *MARC (MACHINE READABLE CATALOGING) and SGML/XML* .
4. Hider, Philip. Metadata or Meta-metadata? Can We Learn from the MARC Experience? *ASIA Library News* 2002 June; Vol. 5 No. 2
5. Lam, K.T. *Moving from MARC to XML* . (Work done in Hong Kong; useful without downloading Chinese character support).
6. Medeiros, Norm. *Making room for MARC in a Dublin Core world* . *Online* 1999 Nov.