

CHUẨN NGHIỆP VỤ CẦN ÁP DỤNG TRONG XỬ LÝ VÀ BIÊN MỤC TÀI LIỆU ĐIỆN TỬ

ThS. Vũ Dương Thuý Nga
Đại học Văn hoá Hà Nội

Giới thiệu một số chuẩn nghiệp vụ cần áp dụng trong xử lý và biên mục tài liệu điện tử: chuẩn mô tả và truy cập tài nguyên, chuẩn áp dụng phương diện kết hợp thuật ngữ chủ đề, chuẩn siêu dữ liệu Dublin Core, chuẩn từ vựng mô tả loại tài liệu, chuẩn ISO mô tả yếu tố ngày tháng và chuẩn ISO mã ngôn ngữ.

Cùng với việc ứng dụng công nghệ thông tin ngày một sâu rộng trong thực tế, các tài liệu điện tử ngày càng trở nên phổ biến trên thế giới và ở Việt Nam. Thực tế đó đã làm các chuẩn biên mục, ngoài một số quen thuộc dành cho các tài liệu truyền thống, thêm đa dạng với sự xuất hiện các chuẩn cho biên mục tài liệu điện tử. Để hướng tới sự chuẩn hoá trong xử lý tài liệu điện tử, một số quy tắc mô tả, các bảng tra cứu chủ đề, chuẩn biên mục mới đã được các thư viện ở nhiều nước trên thế giới quan tâm áp dụng. Bài viết này sẽ đề cập tới một số chuẩn dùng trong biên mục, xử lý tài liệu và một số quy định cần tuân thủ khi áp dụng chúng.

1. Biên mục mô tả: Chuẩn mô tả và truy cập tài nguyên (Resource Description Access- RDA) đã được soạn thảo.

Chuẩn RDA do Ủy ban Hợp tác Phát triển RDA (Joint Steering Committee for Development of RDA- JSC) xây dựng từ năm 2005 và dự kiến ban hành

tháng 11 năm 2009. Mục tiêu của RDA là thay thế AACR2 để trở thành một bộ chuẩn hoàn chỉnh hướng dẫn mô tả và kiểm soát điểm truy cập cho tất cả các dạng tài liệu, hỗ trợ hiệu quả cho quá trình tìm kiếm tài liệu, đặc biệt là trong môi trường điện tử, nơi đòi hỏi độ tương thích cao với công nghệ web và cơ sở dữ liệu. RDA được xây dựng dựa trên mô hình của 4 chuẩn: chuẩn *Yêu cầu chức năng đối với biểu ghi thư mục FRBR* (Functional Requirements for Bibliographic Records); chuẩn *Yêu cầu chức năng đối với dữ liệu kiểm soát FRAD* (Functional Requirements for Authority Data) do Hiệp hội Thư viện Quốc tế IFLA ban hành; chuẩn *Mô hình tổng quát chuẩn siêu dữ liệu Dublin Core* (Dublin Core Metadata Initiatives Abstract Model) và chuẩn *Tương thích (Tương tác) dữ liệu trong hệ thống thương mại điện tử* (interoperability of data in e-commerce systems) do Cộng đồng châu Âu (European Community Infor) ban hành. Về nội dung, RDA đã phản ánh được một số đặc thù của tài

liệu điện tử mà AACR2 chưa thể hiện được.

2. Công cụ định chủ đề tài liệu: bên cạnh bảng Tiêu đề chủ đề của Thư viện Quốc hội Hoa Kỳ, **Chuẩn áp dụng phương diện kết hợp thuật ngữ chủ đề FAST (Faceted Application of Subject Terminology)** - hay còn gọi là *Bảng thuật ngữ chủ đề theo lĩnh vực* (theo cách dịch của Thái Thị Thanh Thủy) đã được biên soạn và triển khai áp dụng.

Chuẩn FAST do Trung tâm Thư viện Máy tính trực tuyến OCLC (Online Computer Library Center) xây dựng từ năm 1998. Nhờ phát triển trên cơ sở bộ đề mục chủ đề LCSH của Thư viện Quốc hội Hoa Kỳ, FAST thừa hưởng những thế mạnh đặc biệt của một bộ chủ đề nổi tiếng nhất thế giới đồng thời đáp ứng các yêu cầu chuyên nghiệp của một bộ đề mục chủ đề trong thời đại thư viện điện tử. FAST được thiết kế theo hướng giảm thiểu việc xây dựng các chuỗi đề mục phức tạp nhưng vẫn duy trì sự phong phú về từ vựng của LCSH. Khác với LCSH các chủ đề của FAST được xây dựng chủ yếu theo phương pháp hậu kết hợp (LCSH thiên về phương pháp tiền kết hợp). Sự đơn giản hoá trong việc kết hợp chủ đề của FAST giúp việc định chủ đề trở nên dễ dàng hơn và người định chủ đề không nhất thiết phải là cán bộ biên mục. FAST đạt được độ tương thích cao với hệ thống tìm kiếm nên rất thuận lợi cho việc tìm kiếm tích hợp các tài liệu thuộc nhiều ngành và nằm ở nhiều

CSDL khác nhau. Một thuận lợi rất lớn đối với các thư viện đang sử dụng LCSH khi chuyển sang sử dụng FAST là hai bộ chuẩn chủ đề này có thể dùng song song với nhau. Nghiên cứu gần đây trên một hệ thống áp dụng cả hai chuẩn này cho thấy sự khác biệt rất nhỏ về kết quả tìm kiếm tài liệu từ hệ thống tích hợp hai chuẩn này. Trong việc phát triển FAST, mục tiêu chính là tạo ra sự tương thích với siêu dữ liệu hiện có, dễ dàng chuyển đổi, hiệu quả, chi phí bảo trì thấp và có khả năng tương tác ngữ nghĩa.

Vốn từ vựng FAST được chia thành tám loại hoặc các khía cạnh riêng biệt: tên cá nhân, tên cơ quan tổ chức, tên địa lý, sự kiện, nhan đề, thời gian, chủ đề và hình thức/thể loại.

Về cấu trúc, FAST bao gồm 4 Hệ đề mục (facets): Hệ đề mục chính (Topical facet), Hệ đề mục địa lý (Geographic facet), Hệ đề mục thể loại (Forms facet) và Hệ đề mục thời kỳ (Periods facet).

Hệ đề mục chính bao gồm các đề mục lớn và các tiểu mục chung tương ứng. Các đề mục của FAST rất giống các đề mục của LCSH, ngoại trừ là các đề mục LCSH thường bao gồm tất cả các tiểu mục chính và các tiểu mục chính này sẽ được tạo lập riêng. Một điểm khác nữa là tất cả các chuỗi đề mục chính của FAST đều nằm trong danh sách kiểm soát (authority file). Các đề mục của FAST được lập theo quy tắc thứ bậc của LCSH nhưng chỉ có các tiểu mục chính mới được ghép với

các đề mục lớn. VD: Quản lý dự án công nghiệp – Xử lý dữ liệu; Đường ống gas tự nhiên – Đặc điểm kinh tế.

Hệ đề mục địa lý bao gồm tất cả các tên địa danh và được thiết lập theo trật tự gián tiếp. VD: Ohio – Columbus thay vì Columbus (Ohio) như trong LCSH. Việc sử dụng quy tắc gián tiếp đã giảm thiểu sự phức tạp của quy tắc kết hợp cả trực tiếp và gián tiếp của LCSH, đồng thời cải thiện giao diện và quy tắc tìm kiếm cho người dùng. Tên địa danh cấp 1 của FAST được giới hạn hơn nhiều so với LCSH do sử dụng bảng Mã Địa danh. VD: Maryland-Quận Worcester [n-us.md]. Tất cả các đề mục địa lý của FAST đều nằm trong danh sách kiểm soát.

Hệ đề mục thể loại bao gồm tất cả các dạng tiểu mục. Đề mục thể loại FAST được thiết lập bằng phương pháp tách tất cả các đề mục thể loại từ các đề mục chính và đề mục địa lý của LCSH. Tất cả các đề mục thể loại của FAST đều nằm trong danh sách kiểm soát.

Hệ đề mục thời kỳ bao gồm tất cả các đề mục thời kỳ cấu trúc theo dạng ngày hoặc giai đoạn. Các đề mục theo niên đại sẽ được phản ánh bằng thời kỳ thực bao quát thời gian đề cập trong nội dung tài liệu. VD: Triều đại Tống, 960-1279 sẽ được thay bởi 960-1279. Tuy nhiên, cấu trúc ngày theo phương thức này của FAST sẽ làm mất thông tin bổ ích, như ở ví dụ này là mất thông tin: *Triều đại Tống* (một triều đại lịch sử nổi tiếng của Trung

Quốc). Do thiết lập đề mục thời kỳ chỉ có một hạn chế khi mô tả là ngày cuối phải lớn hơn ngày đầu theo thứ tự thời gian nên đề mục thời kỳ không nằm trong danh mục kiểm soát như ba hệ đề mục trước đó.

3. Khổ mẫu biên mục: Chuẩn siêu dữ liệu Dublin Core (Dublin Core Metadata Framework) đã nổi lên như một chuẩn biên mục các tài liệu và nguồn tin điện tử được các thư viện ở nhiều nước, trong đó có Việt Nam, quan tâm áp dụng.

Chuẩn siêu dữ liệu Dublin Core (DC) do Trung tâm Thư viện Máy tính trực tuyến OCLC và Trung tâm quốc gia các Chương trình siêu máy tính NCSA (National Centre for Super Computing Applications) xây dựng tháng 3 năm 1995.

Chuẩn Dublin Core hiện tại có hai cấp độ mô tả: đơn giản (simple) và chi tiết (qualified). Ở cấp độ đơn giản, DC bao gồm 15 yếu tố cơ bản: Nhan đề, Tác giả, Chủ đề, Mô tả, Nhà xuất bản, Đồng tác giả, Ngày, Loại tài liệu, Định dạng, Nhận dạng (Định danh), Nguồn, Ngôn ngữ, Quan hệ, Phạm vi và Bản quyền (Right). Ở cấp độ chi tiết, DC bổ sung các phụ tố và các quy tắc (Qualifiers) chuẩn hóa nội dung mô tả nhằm nâng cao chất lượng mô tả và tìm kiếm tài liệu. Các yếu tố bổ sung chỉ nhằm làm rõ yếu tố mô tả hơn là mở rộng hay thay đổi nội dung yếu tố. Với những điểm mạnh này, DC nhanh chóng được chấp nhận và trở thành một

trong những chuẩn được áp dụng rất phổ biến hiện nay trong thư viện điện tử.

Để thực hiện việc biên mục theo DC được thống nhất và đảm bảo chuẩn hoá, một số tiêu chuẩn đã được áp dụng: Chuẩn từ vựng mô tả loại tài liệu, Chuẩn quốc tế mô tả yếu tố ngày tháng (ISO 8601-2004), Chuẩn mã ngôn ngữ (ISO 639-2).

Chuẩn từ vựng mô tả loại tài liệu Dublin Core (DCMI Type Vocabulary): do tổ chức Dublin Core (DCMI) biên soạn năm 2008 nhằm cung cấp một danh sách từ chuẩn thống nhất về loại tài liệu, hỗ trợ việc chuẩn hoá công tác mô tả, nhận dạng và trao đổi dữ liệu giữa các hệ thống và cơ quan thư viện - thông tin. Chuẩn này bao gồm 12 nhóm từ chính: (i) Bộ sưu tập – Collection; (ii) Tập hợp dữ liệu – Dataset; (iii) Sự kiện – Event; (iv) Hình ảnh – Image; (v) Tài liệu tương tác – InteractiveResource; (vi) Hình ảnh động – MovingImage; (vii) Vật thể - PhysicalObject; (viii) Service – Dịch vụ; (ix) Phần mềm – Software; (x) Âm thanh – Sound; (xi) Hình ảnh tĩnh – StillImage; (xii) Văn bản – Text. Mỗi nhóm từ chính bao gồm định nghĩa và tập hợp từ vựng mô tả chi tiết các loại tài liệu thuộc nhóm đó. Ví dụ: nhóm từ Văn bản bao gồm tài liệu dạng từ để đọc như sách, báo, thơ, luận văn, thư từ,... Tổ chức DCMI khuyến nghị các đơn vị thành viên áp dụng chuẩn DCMI Type Vocabulary để mô tả yếu tố Loại tài liệu (Type) trong chuẩn siêu dữ Dublin Core.

Chuẩn ISO mô tả yếu tố ngày tháng (ISO 8601-2004, Data elements and interchange formats – Information interchange – Representation of dates and times): ISO 8601-2004 do Ủy ban Kỹ thuật ISO/TC 154 ban hành năm 2004 và thay thế toàn bộ các chuẩn ISO 8601 trở về trước. Mục tiêu của ISO 8601-2004 là thống nhất định dạng dữ liệu về ngày, tháng và thời kỳ. Ứng dụng của ISO 8601-2004 trong trao đổi thông tin, đặc biệt là giữa các hệ thống xử lý dữ liệu và các thiết bị tương ứng, đã hạn chế tối đa việc mô tả khác nhau trật tự ngày tháng năm, hay giờ phút giây nhằm loại bỏ nhầm lẫn trong nhận dạng số và giảm thiểu chi phí phát sinh. Hệ thống định dạng của ISO 8601-2004 rất linh hoạt cho phép người dùng có thể mô tả đầy đủ thông tin về thời gian hoặc chỉ mô tả ngắn gọn thông tin cần thiết theo nhu cầu. ISO 8601-2004 quy định cách mô tả ngày, thời gian trong ngày, giờ quốc tế, giờ địa phương + chênh giờ, ngày và giờ, thời kỳ và các thời kỳ có tính liên tục. Việc mô tả yếu tố thời gian áp dụng chuẩn ISO 8601-2004 hỗ trợ quá trình nhận dạng trong các hệ thống, tạo khả năng so sánh, sắp xếp và độc lập với ngôn ngữ. Chuẩn ISO 8601-2004 hiện đang được sử dụng rộng rãi trên rất nhiều hệ thống như chương trình máy tính, tài liệu học thuật, tài liệu dạng Web, thương mại điện tử của các cơ quan, tổ chức trên toàn cầu. Tổ chức Dublin Core đã khuyến nghị áp dụng chuẩn ISO 8601-2004 cho yếu tố mô tả ngày (date) trong chuẩn siêu dữ liệu Dublin Core.

Chuẩn mã ngôn ngữ ISO 639-2 (Codes for the representation of names of languages): do các uỷ ban Kỹ thuật ISO TC37, TC46, SC2 và SC4 thuộc Tổ chức Tiêu chuẩn Quốc tế đồng soạn thảo và ban hành năm 1998 để chuẩn hóa việc mô tả ngôn ngữ áp dụng trong lĩnh vực thuật ngữ, từ vựng, thông tin tư liệu (thư viện, dịch vụ thông tin, xuất bản). Bộ chuẩn ISO 639 bao gồm hai phần: 639-1: Alpha-2 code (mã hai ký tự) và 639-2: Alpha-3 code (mã ba ký tự). Chuẩn ISO 639-2 ra đời để hạn chế những khó khăn phát sinh từ việc dùng mã ngôn ngữ 2 ký tự chữ cái của chuẩn ISO 639-1. Chuẩn ISO 639-2 được xây dựng dựa trên danh sách Mã nước của chuẩn MARC nhằm dung hoà sự khác biệt giữa người chuyên dùng thư mục và người dùng quan tâm đến các ứng dụng thuật ngữ. ISO 639-2 hiện đã cập nhật toàn bộ mã ngôn ngữ thuộc bộ 639-1, do đó tất cả các mã ngôn ngữ hai ký tự đã có thêm phần mã ba ký tự tương ứng. Cả hai bộ chuẩn này đều được xây dựng ở dạng mở để

cộng đồng sử dụng có thể cập nhật thêm. Do Tổ chức ISO đã trao quyền kiểm soát đăng ký cho Thư viện Quốc Hội Hoa Kỳ nên chuẩn này sẽ được Thư viện Quốc hội Hoa Kỳ tiếp tục phát triển. Chuẩn ISO 639-2 được tổ chức Dublin Core khuyến nghị áp dụng để mô tả yếu tố ngôn ngữ (language) trong chuẩn Dublin Core.

Các thư viện và cơ quan thông tin ở Việt Nam đang đẩy nhanh tiến trình hiện đại hoá và giành nhiều sự quan tâm cho việc xây dựng và phát triển các thư viện điện tử. Để xây dựng và quản trị các thư viện điện tử không thể bỏ qua công tác biên mục. Chú ý áp dụng các chuẩn nghiệp vụ ngay từ đầu sẽ là một trong những yếu tố và điều kiện tiên quyết giúp cho các thư viện hoạt động hiệu quả, đảm bảo chất lượng, tạo điều kiện cho việc chia sẻ thông tin và phục vụ người dùng tin một cách tốt nhất. Điều đó phần nào sẽ góp phần rút ngắn dần khoảng cách giữa các thư viện Việt Nam với các thư viện tiên tiến trên thế giới.

Tài liệu tham khảo

1. Alan Danskin. *RDA Scope and Structure*, 1 July 2009. <http://www.rda-jsc.org/docs/5rda-scope-rev4.pdf>
2. Dublin Core Metadata Initiatives, 2009. <http://www.dublincore.org>
3. El-Sherbini, M. & Klim, G. *Metadata and cataloging practices. The Electronic Library*, 2004, 22(3): 238 - 248.
4. Gorman, M. *RDA: imminent debacle: The editor of AACR2 takes issue with cataloguing's new direction. (Cataloguing) (Resource Description and Access, revision of the Anglo-American Cataloguing Rules). American Libraries* 2007, 38(11): 64(2).
5. International Standard Organization: ISO 8601-2004, *Data elements and interchange formats – Information interchange – Representation of dates and times*, ISO TC 154. http://www.iso.org/iso/catalogue_detail?csnumber=40874
6. Internet Assigned Numbers Authority. *MIME Media Types. OCLC. FAST Faceted Application of Subject Terminology*, 2009. <http://www.oclc.org/research/projects/fast>
7. Taylor, A. G. *Comparison of keyword searching using FAST vs. Using LCSH*, 2006, ALA Annual Conference. New Orleans, Louisiana (USA) OCLC.
8. <http://fast.oclc.org>