cứ thông tin nào chuyển thành các quyết định, hỗ trợ cho việc quản lý, giải quyết các vấn đề đều được coi là có hiệu quả và như vậy, thực tế có giá trị kinh tế.

Nghiên cứu bối cảnh thực tế ở nhiều nước, có thể thấy nổi lên các vấn đề sau đây:

- Khả năng với tới thông tin hạn chế. Đại biểu các nước phát triển lưu ý rằng, việc không với tới thông tin này là do không biết các phương tiện với tới chứ không phải do việc chi phí cho việc tìm và mua bản quyền quá cao. Có những giải pháp cho tình trạng này, chẳng hạn, các nước công nghiệp hoá xây dựng các catalo kỹ thuật hoặc danh mục (Directory) các xí nghiệp công nghiệp, còn các nước đang phát triển cần ưu tiên ghi lại thông tin do chính mình sản sinh ra, đó là phương tiện trao đổi. Hơn nữa, việc thu hồi và giữ gìn tài sản quốc gia là một mục tiêu KT-XH quan trọng. Ngoài ra, còn có thể tiến hành hợp tác ở qui mô vùng nếu nguồn lực của một nước không đủ làm;

- Ưu tiên đào tạo người dùng tin. Chiến lược phát triển thông tin phải bao gồm:

+ Thành lập nhóm tuyên truyền quan niệm coi thông tin là nguồn lực phát triển kinh tế quốc dân;

+ Đào tạo những người quản lý và cán bộ kỹ thuật để lập kế hoạch xây dựng và phát triển các cơ quan và các dịch vụ thông tin.

- Về mặt pháp lý, những vấn đề sau đây cần được tiếp tục nghiên cứu:

+ Luật quốc tế, nhất là đối với việc tổ chức sở hữu trí tuệ;

+ Thuế quan áp dụng đối với việc cản trở giao lưu thông tin;

+ Những ràng buộc khi mua một sản phẩm được bảo vệ quyền sở hữu công nghiệp.

V. Chủ đề 5: *Chiến lược tổ chức hệ thống và các cơ quan thông tin thích hợp trong quá trình phát triển*

- Hoạt động thông tin cần có một khoản tài chính. Kinh phí cho hoạt động thông tin phải được huy động từ nhiều nguồn khác nhau;

- Cơ chế ngân sách là phương tiện duy trì hoạt động của cơ quan thông tin. Hệ thống dịch vụ phải trả tiền là một phương tiện tốt nhất để đánh giá mức độ thoả mãn người dùng tin và đo gần đúng kết quả hoạt động của một chương trình. Nhiều đại biểu cho biết, kinh nghiệm cho thấy, 70% chi phí dịch vụ thông tin nghề nghiệp quốc gia là do tiền nhận được từ phía người dùng tin;

- Kiến nghị các trung tâm phân tích tin nên dựa vào các cơ quan lưu trữ quốc gia. Tuy nhiên, hiện nay cách phân loại trong các hệ thống thông tin và các cơ quan lưu trữ có khác nhau, điều đó có thể cản trở việc hợp tác kỹ thuật hai bên. Cũng có ý kiến đề cập đến việc giúp đỡ nước nghèo bảo tồn các tài liệu lưu trữ cho nước mình vì đó là nguồn thông tin rất quan trọng đối với việc phát triển của một quốc gia.

Việt Anh

(Tổng hợp từ các nguồn tin từ UNESCO, PGI).

Biên mục tương lai với Dublin Core, SGML DTD, MARC DTD, DOI và TEI

Trong quá trình nghiên cứu tài liệu về các tiêu chuẩn và khổ mẫu mới, đặc biệt trong thời kỳ phát triển của công nghệ WWW, thì đã có sự tranh luận về những nhược điểm và sự lạc hậu của MARC so với Dublin Core và những khổ mẫu tương lai kết hợp giữa MARC và DTD (Document Type Definition, Cấu trúc định dạng tư liệu)

như MARC/DTD. Người ta đã bàn đến cả các chương trình chuyển đổi sang MARC/ DTD.

Khổ mẫu tiêu chuẩn siêu dữ liệu Dublin Core là một tập hợp các yếu tố đơn giản nhưng hữu hiệu trong việc mô tả một loạt nguồn tin trên mạng. Dublin Core gồm có 15 yếu tố, mà ngữ nghĩa của chúng đã được xác lập với sự đồng thuận của một nhóm chuyên gia liên ngành quốc tế thuộc các lĩnh vực thư viện học, tin học, mã hoá văn bản, bảo tàng và các lĩnh vực liên quan. Mọi yếu tố đều không bắt buộc và có thể lặp, ngoài ra còn có một số lượng hạn chế các từ hạn định và định ngữ có thể sử dụng để tiếp tục tinh chỉnh (chứ không phải mở rộng) ý nghĩa của các yếu tố.

Mặc dầu Dublin Core thích hợp với các đối tượng kiểu tư liệu, nó có thể được áp dụng cho các nguồn tin khác. Việc sử dụng Dublin Core cho các nguồn tin phi tư liệu phụ thuộc một phần vào mức độ các siêu dữ liệu của các nguồn đó giống các siêu dữ liệu tư liệu điển hình và vào mục đích mà siêu dữ liệu của các nguồn này phục vụ.

Dublin Core có những đặc điểm sau: Tạo lập và duy trì dễ dàng; ngữ nghĩa dễ hiểu; phạm vi sử dụng quốc tế rộng lớn với các phiên bản đa ngôn ngữ; khả năng mở rộng các yếu tố thuận tiện.

Thuật ngữ "MARC DTD"có liên quan đến việc áp dụng ngôn ngữ đánh dấu tổng quát hoá chuẩn (SGML). SGML là kỹ thuật trình bày tư liệu dưới dạng đọc máy đã được thông qua như một tiêu chuẩn quốc tế: ISO 8879 (Xử lý thông tin- Văn bản và các hệ thống văn phòng- Ngôn ngữ đánh dấu tổng quát hoá chuẩn). Tiêu chuẩn này được phát triển để thoả mãn nhu cầu cần có một tiêu chuẩn chung (không thuộc quyền sở hữu của riêng ai) để mã hoá văn bản sao cho các dữ liệu máy đọc có thể trao đổi được giữa các môi trường mã hoá văn bản khác nhau. SGML được dùng rộng rãi trong công nghiệp xuất bản nơi sử dụng các hệ tin học khác nhau để tạo ra các tư liệu. SGML hỗ trợ cho việc xác định các tập hợp yếu tố (kể cả bài tóm tắt) hình thành các dạng tư liệu riêng (thí dụ, các bài báo chí). MARC DTD xử lý các biểu ghi biên mục đọc máy như là một dạng tư liệu riêng biệt. Cấu trúc dữ liệu này xác định tất cả các yếu tố có thể hình thành biểu ghi MARC song song với danh mục các yếu tố dữ liệu đã được xác định trong 5 khổ mẫu USMARC: Khổ mẫu thư mục, khổ mẫu tiêu đề/đề mục chuẩn, khổ mẫu về vốn tư liệu, khổ mẫu phân loại, khổ mẫu thông tin cộng đồng)

Mục đích ban đầu của Dự án MARC DTD là tạo lập các cấu trúc định dạng tư liệu theo ngôn ngữ đánh dấu tổng quát hoá chuẩn để hỗ trợ cho việc chuyển đổi dữ liệu biên mục từ cấu trúc dữ liệu của MARC sang SGML (và ngược lại) mà không bị mất dữ liệu. Mặc dầu cả hai tiêu chuẩn ISO 2709 và ISO 8879 đều cung cấp các kỹ thuật chuẩn để mã hoá dữ liệu, nhưng nếu như không phát triển một MARC DTD chuẩn (hoặc một tập hợp của các cấu trúc định dạng tư liệu) thì mối quan hệ giữa hai tiêu chuẩn này có thể được xác lập theo những cách khác nhau. Động lực của Dự án là nguyện vọng muốn chuyển đổi bằng máy một cách chuẩn mực không vi phạm quyền sở hữu trí tuệ, giữa các dữ liệu mã hoá theo MARC và SGML. Dự án có 2 nhiệm vụ chính: 1) Phát triển các cấu trúc định dạng tư liệu theo ngôn ngữ đánh dấu tổng quát hoá chuẩn (SGML DTD) tương ứng với 5 khổ mẫu USMARC; và 2) Phát triển các phần mềm tiện ích có khả năng chuyển đổi giữa hai tiêu chuẩn mã hoá. Văn phòng phát triển mạng và tiêu chuẩn MARC thuộc Thư viện quốc hội Mỹ đã đảm nhiệm cả hai nhiệm vụ này.

4 nguyên tắc của việc thiết kế MARC/ DTD là:

- Cấu trúc này phải có tính

phổ quát đối với các ứng dụng của MARC và SGML để có thể trao đổi lẫn nhau được;

- Có khả năng chuyển đổi mà không làm mất nội dung trí tuệ hay thông tin liên quan đến những yếu tố cơ bản của cấu trúc biểu ghi khác, nghĩa là giữ nguyên nhãn và ngữ nghĩa của MARC trong khi chuyển đổi sang SGML;

- Tính linh hoạt, nghĩa là tăng cường khả năng tuỳ chọn và tính lặp;

- Tính thân thiện với người dùng (có thể sử dụng DTD với các hệ máy tính không chuyên dụng); có quan hệ với TEI (Tiêu chuẩn mã hoá và trao đổi văn bản điện tử): sao cho một yếu tố trong đầu tệp (file header) TEI MARC DTD có thể liên kết với MARC DTD.

Nhu cầu tạo ra một khung mã hoá chung cho các cấu trúc văn bản phức hợp nhằm làm giảm sự đa dạng trong thực tiễn mã hoá hiện nay, đơn giản hoá việc xử lý bằng máy và khuyến khích sự chia sẻ các văn bản điện tử là nguyên nhân dẫn đến việc thiết lập Dự án hợp tác mang tính chất quốc tế TEI có mục đích phát triển các tài liệu hướng dẫn cho việc biên soạn và trao đổi các văn bản điện tử phục vụ cho nghiên cứu khoa học.

Đầu tệp TEI, một bộ phận bắt buộc phải có của các văn bản tuân thủ sáng kiến mã hoá văn bản (TEI), có 4 phần chính:

- Phần mô tả tệp, chứa mô tả thư mục đầy đủ một tệp tin học kết hợp với thông tin về nguồn hoặc những nguồn mà từ đó tạo ra văn bản điện tử;

- Phần mô tả cách mã hoá, phản ánh quan hệ giữa văn bản điện tử và nguồn gốc của nó: phần này cho phép mô tả chi tiết cách thức văn bản được chuẩn hoá trong quá trình chuyển tả, cách thức người mã hoá phân giải các trường hợp không rõ ràng trong nguồn,...;

- Phần mô tả diện văn bản (Text profile), chứa các thông tin về văn bản theo khía cạnh phân loại và văn cảnh như chủ đề, tình huống sản sinh ra văn bản, các nhân vật được mô tả trong văn bản hoặc tham gia vào việc tạo lập văn bản,...;

- Phần lịch sử sửa đổi văn bản, cho phép người mã hoá cung cấp những dữ kiện lịch sử về những thay đổi trong quá trình phát triển văn bản điện tử.

Tài liệu hướng dẫn đầu tệp TEI còn cung cấp những thông tin về mối quan hệ của đầu tệp với biểu ghi MARC. Mục đích của TEI trong việc tạo lập đầu tệp là để đảm bảo thông tin được yêu cầu cho một biểu ghi biên mục đều có thể tìm được từ một đầu tệp TEI và hơn thế nữa còn đảm bảo cho việc ánh xạ từ đối tượng này sang đối tượng kia càng đơn giản và càng trực tiếp thì càng tốt.

Tuy nhiên, tài liệu hướng dẫn cũng nói rằng khác biệt quan trọng nhất giữa biểu ghi MARC và đầu tệp TEI là chức năng của các đối tượng này: Về cơ bản, biểu ghi MARC vẫn còn là một phiên bản điện tử của tờ phiếu mục lục với những hạn chế của mô hình phiếu. Trong khi đó, đầu tệp TEI chẳng những cung cấp thông tin thư mục đầy đủ (như một tờ phiếu mục lục) mà còn cung cấp cả những thông tin tư liệu phi thư mục hỗ trợ cho người hoặc máy phân tích văn bản điện tử đã được đầu tệp tư liệu hoá. Phần lớn thông tin phân tích này nằm trong phần mô tả diện văn bản, mô tả cách mã hoá và lịch sử sửa đổi, mà biểu ghi MARC hầu như không cung cấp trực tiếp, và nếu có thì được ghi vào khối trường phụ chú không cấu trúc hoá (5XX).

Trong khi nghiên cứu về biên mục hiện đại với môi trường số, người ta quan tâm nhiều đến việc phát triển MARC DTD vì kỹ thuật SGML và Internet ngày càng phát triển. Người ta cũng đề cập tới yếu tố nhận dạng đối tượng số (DOI: Digital Object Identifier), một chỉ số nhận dạng tư liệu số hoá có chức năng gần giống như ISBN,

ISSN trong môi trường xuất bản và mô tả ấn phẩm. Thực ra, cũng đã có một chỉ số nhận dạng nội dung được dùng phổ biến trên mạng Internet, đó là yếu tố định vị nguồn đồng nhất (URL: Uniform Resource Locator) dùng để tìm những trang đặc biệt trên mạng toàn cầu WWW. Tuy nhiên, URL cũng chưa thoả mãn cho việc nhận dạng nội dung, mà chỉ xác định vị trí. Chính DOI mới thực là yếu tố nhận dạng nội dung một cách đơn trị và lâu dài. DOI có thể được mô tả như là một chỉ số nhận dạng "cố kết"(persistent) của các thực thể có quyền sở hữu trí tuệ, với những biểu hiện hữu hình (sách, báo chí, CD, băng hình, ...) cũng như những biểu hiện vô hình (các tệp tin số hoá) và cả mức độ thực hiện quyền sở hữu trí tuệ hoặc nội dung trừu tượng tiềm ẩn trong những biểu hiện khác nhau. DOI là một chỉ số cố kết bởi vì ngay cả khi chuyển nhượng quyền sở hữu trí tuệ thì việc nhận dạng thực thể (tư liệu số) không hề thay đổi. DOI có liên quan đến công nghiệp nội dung. Từ 1997, Cơ quan DOI quốc tế đã bắt đầu quá trình phát triển và cải tiến hệ thống chỉ số nhận dạng này về mặt công nghệ, qui trình và chính sách. Như vậy, DOI không chỉ là một phương tiện định danh các đối tượng số, nó là một hệ thống tích hợp. Hệ thống DOI được hình thành từ một số thành tố tương tác, phụ thuộc vào nhau để tạo nên giá trị chung của chúng. Hệ thống DOI có 4 thành tố như sau:

- Liệt kê: Gán 1 số (đúng hơn là, một chuỗi chữ- số, vì có khi sử dụng cả chữ cái) cho một thực thể có quyền sở hữu trí tuệ;

- Mô tả: Tạo ra một bản mô tả (siêu dữ liệu) của thực thể được nhận dạng bằng DOI.

- Phân giải: Làm cho chỉ số nhận dạng này có hiệu lực bằng cách cung cấp thông tin về những gì mà DOI phải phân giải và công nghệ cung cấp các dịch vụ cho người dùng.

- Chính sách: Các qui tắc chi phối hoạt động của hệ thống.

Cho đến nay, trên toàn thế giới đã có hơn 3 triệu chỉ số được phân phối bởi hơn 100 tổ chức thực hiện việc đăng ký.

Các khổ mẫu, tiêu chuẩn và chỉ số hiện đại kể trên mặc dầu có chú trọng tới việc trao đổi các nguồn tin điện tử và phản ánh nội dung số của tư liệu được biên mục, còn đang trên đường hoàn thiện trong một môi trường phát triển chưa ổn định được như MARC, tuy nhiên, chúng ta cần tiếp tục nghiên cứu cho tương lai lâu dài.

Vũ Văn Sơn

ĐÀO TẠO CÁN BỘ CHUYÊN NGÀNH THÔNG TIN Ở THẾ KỶ 21

Có thể nói rằng, xã hội hiện nay gần như ở trong giai đoạn thoát ra khỏi quá khứ để bước vào một tương lai có rất nhiều những điều bất định. ởvào những thời điểm như vậy thường có những thay đổi và nhiều vấn đề được đặt ra. Tương tự, tương lai của ngành thông tin cũng không còn là những vấn đề mà mọi người đã từng quen thuộc.

Cuộc cách mạng máy tính và viễn thông và những điều diễn ra trong hai thập kỷ vừa qua đã có tác động mạnh tới ngành thông tin, trong đó các cán bộ thông tin, thư viện và lưu trữ chiếm vị trí trung tâm. Nếu như