
Xây dựng thư viện điện tử và vấn đề số hoá tài liệu ở Việt Nam

ThS. Nguyễn Tiến Đức

Trung tâm Thông tin KHCN Quốc gia

Tóm tắt: Trình bày tiếp cận xây dựng thư viện điện tử (TVĐT). Xem xét khía cạnh cấu trúc, hạ tầng cơ sở kỹ thuật và phát triển kho tư liệu số hoá của TVĐT. Đề cập việc tổ chức số hoá tài liệu trong phạm vi Mạng lưới các tổ chức thông tin KHCN ở Việt Nam.

Phát triển từ thư viện truyền thống thành TVĐT đang là xu hướng tất yếu ở tất cả các nước. Để xây dựng được một TVĐT theo đúng nghĩa, cần có một số quan điểm thống nhất, có cách tiếp cận đúng và lựa chọn những bước đi thích hợp. Trong bài báo này, chúng tôi luận bàn điều kiện để xây dựng TVĐT ở Việt Nam.

I. Tiếp cận xây dựng Thư viện điện tử

Để xây dựng TVĐT, chúng ta cần quan tâm nhiều vấn đề mà nổi bật là 4 khía cạnh chủ yếu: Cấu trúc của TVĐT; Hạ tầng cơ sở kỹ thuật; Kho tư liệu số hoá; Các vấn đề bảo quản, khai thác và bản quyền.

Dưới đây sẽ đi qua vấn đề 1, 2 và sau đó tập trung vào vấn đề thứ 3.

Cấu trúc của Thư viện điện tử

Các TVĐT đều được bố trí trên “Giao diện web”, trên đó, ngoài những vùng chung như: Giới thiệu về cơ quan, về hệ thống, về thư viện; Hướng dẫn sử dụng và các công cụ trợ giúp thì phần chủ yếu là nội dung, tức là “Tài nguyên thông tin”.

Phần thứ nhất của “Tài nguyên thông tin” thông thường là Danh mục chủ đề/Directory. Phần này được cấu trúc theo trình tự từ chung đến riêng, từ tổng quát đến chi tiết, từ ngoài vào trong theo thứ bậc, ví dụ: Mạng/vấn đề lớn; Tiếp đó là các mục/vấn đề nhỏ; Mỗi mục này lại chia nhỏ dần theo cấu trúc hình cây: Cây – cành; Cành – nhánh to; Nhánh to – nhánh nhỏ; Nhánh nhỏ – nhánh nhỏ hơn... Cùng với các phân chia này là các điểm liên kết: đảm bảo các mối quan hệ nhiều chiều giữa các nhánh cũng như giữa nhánh với các cành. Cách tổ chức như vậy nhằm tạo thuận tiện cho người dùng trong khai thác thông tin. Thông thường trong TVĐT các xuất bản phẩm điện tử (tạp chí, bản tin, kỷ yếu ...) được bố trí sắp xếp theo kiểu này và để tìm kiếm tài liệu theo chủ đề ta có thể “Click-Nháy” vào đề mục tương ứng tuần tự từ ngoài vào trong. Nếu muốn tìm/truy cập nhanh tới tài liệu cần phải có sự hỗ trợ của Máy tìm tin/Search engine thông qua các Lệnh tìm cụ thể,....

Trong TVĐT còn có phần “Tài nguyên” thứ hai - quan trọng hơn, đó là các tổ hợp CSDL, biểu hiện Danh mục các CSDL, thường được sắp xếp theo chủ đề hoặc theo văn

chữ cái. Người dùng có thể tiếp cận tới các CSDL này để khai thác thông tin theo các cấp độ khác nhau: từ thư mục tới toàn văn; khai thác riêng rẽ từng CSDL hay khai thác theo Nhóm CSDL,... Mức độ khai thác đến đâu tùy thuộc vào khả năng của hệ thống và đặc biệt là sự cho phép của cơ quan chủ quản, các lệ phí tương ứng.

Thành phần thứ 3 trong TVĐT là phần *Liên kết* tới các nguồn tài nguyên thông tin bên ngoài. Đây là thế mạnh của TVĐT. Tuy nhiên, mức độ và khả năng liên kết đến đâu phụ thuộc vào sự hợp tác với các cơ quan khác và việc khai thác các tầng thông tin số hoá đó cũng có những khác biệt. Chẳng hạn: có vùng/mảng thông tin được khai thác tự do, miễn phí, nhưng có những vùng/CSDL phải có mật khẩu, phải trả tiền,....

Như vậy, TVĐT không chỉ có một hệ thống mà có thể gồm nhiều hệ khác nhau. Tuy nhiên, các tài nguyên thông tin, các CSDL đó liên kết được với nhau trong một chế độ phục vụ thống nhất. Tức là khi được yêu cầu, chúng xuất hiện đối với người dùng như thể chúng cùng trong một hệ thống. Để tích hợp được như vậy, chúng ta phải có các chương trình phần mềm hỗ trợ, phải áp dụng các chuẩn nhất định trong xử lý, trong quản trị, trong trao đổi dữ liệu cũng như phải có các công cụ tìm kiếm (search engine), chuyển tải, lưu trữ thông tin,....

Tóm lại, cấu trúc của TVĐT thực chất là cấu trúc của một Trang Web có liên kết đến các nguồn tin số hoá, trong đó quan trọng nhất là các CSDL toàn văn, được tổ chức theo cấu trúc có khả năng đáp ứng cho việc khai thác qua chế độ mạng on-line.

Vấn đề kỹ thuật, hạ tầng cơ sở và phần mềm TVĐT

Một TVĐT phải có hạ tầng cơ sở đủ mạnh, đó là:

- Mạng Intranet có tốc độ kết nối nhanh với INTERNET;
- Hệ thống máy chủ lớn thực hiện việc quản trị và các dịch vụ khác nhau: Máy chủ Web, Máy chủ FPT, Mail, các Máy chủ lưu, bảo trì dữ liệu; Máy chủ Firewall, Máy chủ cho các ứng dụng khác,....
- Hệ thống máy trạm để cập nhật, khai thác thông tin;
- Các thiết bị công nghệ chuyên dụng cho TVĐT: mã vạch, quản lý và in thẻ, máy quét, máy sao CD,....

Về phần mềm: Đến nay, trên thế giới có nhiều phần mềm phục vụ cho việc xây dựng và phát triển TVĐT. Mỗi phần mềm đều có những ưu, nhược điểm riêng nhưng thông thường một phần mềm khả dĩ phải có các module chính của thư viện, như: Bổ sung; Biên mục; Quản lý Kho; Phục vụ bạn đọc; Mục lục trực tuyến; Phân hệ lưu hành; Quản lý tài liệu điện tử; Truy hồi và trình bày thông tin; Mượn liên thư viện; Quản trị hệ thống.

Ở Việt Nam, trong 5-7 năm gần đây đã xây dựng, phát triển phần mềm thư viện, như: **Libol** (CTy Tinh Vân), **Ilib** (Cty CMC), **Vebrary** (Cty Lạc Việt), **Elib** (Cty VNNetsoft). Qua đánh giá, hiện tại 02 phần mềm Libol và Ilib hoàn toàn có thể đáp ứng được chức năng của một TVĐT hiện đại ở Việt Nam với các cấp độ khác nhau (đến hết năm 2004, mỗi phần mềm này đã được vài chục thư viện ứng dụng).

Ngoài ra, để tổ chức TVĐT ta cũng cần có: Phần mềm Hệ thống: Hệ điều hành và Hệ quản trị các CSDL; Phần mềm xuất bản điện tử, xuất bản CD-ROM.

Theo ước tính, để trang bị các thiết bị nêu trên cũng như có phần mềm TVĐT và tạo lập nội dung TVĐT ban đầu ta phải đầu tư khoảng 5-7 tỷ VND, thậm chí một số người cho rằng ít nhất phải đầu tư 1 triệu USD (16 tỷ VND) cho một TVĐT trung bình của một ngành/lĩnh vực.

II. Tạo lập và phát triển Kho tư liệu số hoá- vấn đề trọng tâm trong xây dựng TVĐT

Phần quan trọng nhất trong TVĐT chính là Kho tư liệu số hoá của bản thân cơ quan thông tin/thư viện chủ quản. Có 3 cách để tạo lập Kho này, đó là:

- Tự tiến hành số hoá nguồn tư liệu trên giấy của Thư viện. Tức là chuyển tài liệu hiện có sang dạng số bằng phương pháp quét hay nhập lại thông tin từ bàn phím... Đây là hướng phải đầu tư lớn, đầu tư liên tục và tốn kém thời gian, tiền của, công sức;
- Bổ sung/tích hợp nguồn tin điện tử thông qua việc mua, trao đổi tài liệu điện tử đang được xuất bản (bản tin, tạp chí điện tử, các chế bản điện tử trước khi in ra trên giấy). Chúng ta đều biết: hầu hết các ấn phẩm hiện nay đều vừa xuất bản trên giấy vừa tồn tại dưới dạng điện tử và nếu tận dụng được nguồn này, ta sẽ tiết kiệm được nhiều công sức, thời gian;
- Xây dựng các liên kết (tạo khả năng truy cập) đến các nguồn tài liệu trên INTERNET, nhất là nguồn của các cơ quan có cùng diện chuyên đề bao quát.

Tạo lập và phát triển Kho tài liệu số của riêng mỗi cơ quan thông tin/thư viện là vấn đề lớn nhất trong xây dựng TVĐT. Công việc này đòi hỏi phải đầu tư lớn và liên tục. Để làm tốt công việc này, các cơ quan xây dựng TVĐT cần có cách tiếp cận hợp lý, khả thi và kinh tế. Cụ thể là:

1. Nhất thiết phải lập kế hoạch sát sao và ưu tiên đầu tư cho việc thu thập, xử lý và số hoá nguồn tin cơ bản, nguồn tin tiềm năng của riêng mình. Coi đây là nhiệm vụ trọng tâm, thậm chí phải triển khai mạnh trước khi bắt tay vào xây dựng TVĐT.

Nếu không có sự đi trước này, khi ta xây dựng xong hạ tầng mạng và có các phần mềm Hệ thống, phần mềm TVĐT đầy đủ nhưng đến lúc đó cơ quan vẫn không có hoặc có rất ít tài liệu số hoá của bản thân chắc chắn TVĐT đó không thể phát huy được hiệu quả; và như vậy, không tương xứng với kinh phí đầu tư nhiều tỷ đồng của Nhà nước.

Trong việc lựa chọn tài liệu để số hoá, ta phải ưu tiên các tài liệu đặc thù của thư viện, các tài liệu duy nhất và có giá trị lâu dài để trao đổi, ví dụ: các tài liệu quý hiếm, các sưu tập có giá trị, không ở đâu có...; ưu tiên số hoá trước hết đối với tài liệu chưa ở đâu số hoá, tài liệu tiếng hiếm, tiếng Việt,... Song song với việc số hoá là việc xây dựng các Siêu dữ liệu đối với từng tài liệu và cập nhật tài liệu đã được số hoá này vào CSDL tương ứng để phục vụ kịp thời cũng như làm cơ sở cho việc xây dựng TVĐT sau này.

Ngoài ra, ta cũng cần quan tâm đến chất lượng của việc số hoá tài liệu cũng như cần sao lưu đầy đủ, kịp thời các tài liệu số hoá đó để tránh rủi ro cũng như tránh phải làm đi làm lại (lãng phí công sức, tiền của). Điều này phụ thuộc nhiều vào công tác tổ chức, phụ thuộc vào cán bộ thực hiện cũng như thiết bị và quy trình số hoá.

2. Phải phối hợp, tận dụng sản phẩm số hoá của các cơ quan thông tin/thư viện khác, nhất là của những cơ quan có cùng diện chuyên đề bao quát. Trên cơ sở đó mới có thể tăng nhanh “nguồn tin” của mình, tiết kiệm được thời gian, công sức và tiền của.

III. Tổ chức số hoá tài liệu trong phạm vi quốc gia và tiến tới Nền Công nghiệp nội dung

Trong phạm vi quốc gia, Nhà nước cần có chiến lược đầu tư và triển khai kế hoạch tổng thể trong việc số hoá đối với các nguồn tin quan trọng, có ý nghĩa lâu dài.

Xác lập chính sách ưu tiên số hoá

Theo chúng tôi, Nhà nước ta nên có chính sách ưu tiên hỗ trợ cho việc số hoá đối với các tài liệu sau:

Về dạng tài liệu:

- Văn bản quy phạm pháp luật;
- Báo cáo kết quả các nhiệm vụ KHCN (các đề tài/đề án các cấp);
- Tài liệu điều tra cơ bản, quy hoạch, bản đồ;
- Luận văn sau đại học;
- Tài liệu hội nghị, hội thảo khoa học;
- Tài liệu tiêu chuẩn, quy phạm; Tài liệu sở hữu công nghiệp.

Về ngôn ngữ: ưu tiên số hoá tài liệu Hán nôm, tài liệu tiếng Việt. Hạn chế số hoá tài liệu tiếng Anh.

Về lĩnh vực: ưu tiên số hoá tài liệu KHCN những ngành mũi nhọn.

Xác lập và hình thành một mạng lưới các tổ chức số hoá tài liệu

Trong mạng lưới các cơ quan thông tin, thư viện cần có sự phân công, phối hợp giữa các cơ quan trong việc số hoá tài liệu.

Nguyên tắc đầu tiên trong phân công là cơ quan nào có kho tư liệu chuyên môn hoá với số lượng tài liệu nhiều nhất về lĩnh vực nào thì sẽ đảm nhận số hoá nguồn tư liệu về lĩnh vực đó và cơ quan này sẽ được Nhà nước hỗ trợ trong việc số hoá.

Trên cơ sở phân công như vậy, ta xác lập được *một mạng lưới các tổ chức số hoá tài liệu* trên quy mô lớn, hiện đại, bao quát hầu hết những dạng tài liệu, những nguồn tin quan trọng nhất, những tài liệu có giá trị lâu dài ở tầm quốc gia. Từ đó hình thành mạng trao đổi thông tin toàn quốc và tạo điều kiện thuận lợi cho các cơ quan TT-TV trọng điểm trong xây dựng và phát triển TVĐT của mình.

Nghiên cứu, lựa chọn và xác lập những chuẩn, những quy định chung trong việc số hoá tài liệu

Trong mạng lưới các cơ quan TT-TV tiến hành số hoá cần có những chuẩn cũng như các quy định thống nhất đối với việc số hoá tài liệu. Những chuẩn này được xác định trên cơ sở nghiên cứu áp dụng chuẩn của nước ngoài vào điều kiện cụ thể của Việt Nam. Mỗi cơ quan tham gia mạng lưới số hoá đều phải tuân thủ nghiêm ngặt những chuẩn này sao cho các tài liệu số hoá luôn đảm bảo về mặt chất lượng cũng như về tiến độ thời gian;

đồng thời chúng được tổ chức trong các CSDL có cấu trúc tương hợp hoặc dễ dàng trong chuyển đổi. Có như vậy, các tài liệu số hoá đó mới có thể phục vụ rộng rãi trong toàn mạng lưới và mang lại hiệu quả cao.

Tổ chức số hoá tài liệu ở quy mô công nghiệp

Hoạt động số hoá ở quy mô lớn được gọi là “Ngành công nghiệp nội dung” (CNND). Theo định nghĩa khái quát người ta coi CNND bao gồm: thu thập thông tin; phân tích nội dung để phân loại, lưu trong bộ nhớ hoặc đưa lên mạng máy tính; bao gói thông tin thành các CSDL (trên CD, DVD, ổ cứng –HDD, thiết bị lưu trữ di động (HDD-RAM), nhân sao và cung cấp/bán các CSDL đó.

Nhiều người cho rằng, CNND là một nhánh của CNTT, vì nó gắn với CNTT. Sản phẩm của CNND là sản phẩm của CNTT, từ công nghiệp phần mềm, thu thập thông tin, xử lý, lưu trữ thông tin đến cung cấp thông tin. CNND không chỉ liên quan đến phần cứng, phần mềm mà còn liên quan đến các lĩnh vực khác như lưu trữ, sản xuất các thiết bị lưu trữ (vật mang tin hiện đại).

Như vậy, nguyên vật liệu đầu vào của ngành CNND là thông tin và 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 bán/cung cấp cho những đối tượng cụ thể nhằm phục vụ cho một/một số hoạt động KT-XH nhất định.

Tuy nhiên, để xây dựng được một nền CNND ở tầm quốc gia cần phải:

- Có một hành lang pháp lý đủ mạnh để: một mặt, tạo điều kiện thuận lợi cho các tổ chức, cá nhân tìm kiếm, xử lý, số hoá, bao gói, chuyển giao thông tin; mặt khác, bảo vệ thiết thực bản quyền tác giả đối với các CSDL cũng như những thông tin mà các cá nhân, tổ chức sở hữu và tất nhiên là bảo đảm bí mật quốc gia;
- Có sự phối hợp chặt chẽ về mặt tổ chức, liên kết, phân công phân mảng giữa các cơ quan có hoạt động số hoá tài liệu;
- Có lực lượng cán bộ chuyên trách thu thập, số hoá, bao gói thông tin (lực lượng này không phải chỉ là cán bộ thông tin, thư viện);
- Trang bị các thiết bị hiện đại cho các cơ sở tham gia số hoá: ví dụ các máy quét chuyên dụng (nhanh, chất lượng cao, quét được các khổ lớn, màu sắc đẹp...); các USB, các máy chủ sao lưu với dung lượng lớn (hàng nghìn GB); các máy sao CD chuyên dụng,...

Chắc chắn rằng, nếu tổ chức tốt ở Việt Nam ta, các sản phẩm số hoá này sẽ có thị trường không phải chỉ trong nước mà cả ở nước ngoài. Nền CNND sẽ hỗ trợ, cung cấp các nguồn tin số hoá cho các TVĐT và làm cho hoạt động thông tin KHCN sôi động, hấp dẫn hơn, cán bộ thông tin sẽ có nhiều việc làm và thu nhập khá hơn.

Kết luận

Tiến tới TVĐT là xu hướng tất yếu, là mong muốn của mỗi thư viện. Tuy nhiên, để có được một TVĐT hoạt động có hiệu quả, phát huy được thế mạnh “*thông tin đặc thù*” của mình, cơ quan chủ quản cần có kế hoạch sát thực, lựa chọn bước đi phù hợp. Trong xây dựng và phát triển TVĐT, việc tạo lập Kho tư liệu số hoá là nhiệm vụ hàng

đầu. Giải quyết nhiệm vụ này đòi hỏi cơ quan chủ quản (cơ quan thông tin, thư viện) phải có chương trình thu thập, số hoá tài liệu và tạo lập các CSDL một cách đầy đủ, kịp thời, đặc biệt là các CSDL toàn văn với các tài liệu có giá trị lâu dài thuộc phạm vi bao quát của cơ quan.

Tài liệu tham khảo

1. What are digital libraries?/Donald J. Waters// CLIR, No 4, July/August 1999. URL: <http://WWW.clir.org/pubs/issues/issue04.html>.
2. Digital libraries: Definitions, issues and challenges / Gary Cleveland // UDT Occasional paper #8. URL:
3. NII, GII and 12, and IT*2 Initiatives: Implications to the digital Library development in the US / Ching-Chih Chen (edited). IT and global digital Library development. Massachsets : MicroUse information, 1999, pp 49-64.
4. European development in digital libraries / David Raitt // Ching-Chih Chen (edited). IT and global digital Library development. Massachsets: MicroUse information, 1999, pp 345-356.
5. Cao Minh Kiểm. Thư viện số: Định nghĩa và vấn đề.-Tập chí Thông tin & Tư liệu, 2000, số 3, tr. 5- 11.
6. Tạ Bá Hùng. Phát triển nội dung số ở Việt Nam: những nguyên tắc chỉ đạo .-Tập chí Thông tin & Tư liệu, 2000, số 1, tr. 2-6.

Building up electronic libraries and digitization of documents in Vietnam

Nguyen Tien Duc

Journal of Information and Documentation, 2005, no. 2, pp..

Abstracts: Deals with the approach to building up electronic libraries; Considers structural aspects, technical infrastructure and digitized collections of electronic libraries; Mentions organization of digitizing documents within information institutions in the National Information System.