

MỘT SỐ VẤN ĐỀ VỀ SỐ HÓA TÀI LIỆU

Dương Duy Tiến – Cán bộ Thư viện tỉnh Nghệ An

I. ĐẶT VẤN ĐỀ

Hiện nay, sự phát triển nhanh chóng của kỹ thuật công nghệ thông tin đã hỗ trợ đắc lực cho nhu cầu truyền tin và khai thác thông tin dữ liệu của nhân loại. Trong đó, hệ thống thông tin điện tử Internet được mọi người quan tâm sử dụng nhiều nhất nhờ tính thông dụng trong việc giao tiếp, tìm kiếm và đọc thông tin bằng tính năng ưu việt của các phần mềm ứng dụng về quản lý, số hóa tài liệu.

Trên hệ thống thông tin điện tử Internet không chỉ diễn ra sự giao tiếp, phổ cập thông tin mà còn hình thành nên một thị trường thông tin hàng hóa rộng lớn, nhiều tiềm năng. Bên cạnh đó, do yêu cầu của quá trình hội nhập đang diễn ra mạnh mẽ, việc công khai minh bạch các văn bản, chính sách của nhà nước đến mọi đối tượng trở thành vấn đề bắt buộc; Đặc biệt việc nâng cao chất lượng thông tin phục vụ nhu cầu nghiên cứu quyết định đầu tư, sản xuất cho các nhà đầu tư trong và ngoài nước trên lĩnh vực quản lý nhà nước, kinh tế, văn hóa, xã hội, lịch sử truyền thống cần phải được chú trọng; Một trong những nguồn thông tin được các nhà đầu tư quan tâm nhiều nhất là thông tin từ tài liệu lưu trữ.

Mặt khác dưới góc độ quản lý nhà nước, việc bảo quản và sử dụng có hiệu quả giá trị tài liệu lưu trữ có ý nghĩa cực kỳ quan trọng đối với kinh tế, xã hội, lịch sử của mỗi quốc gia và địa phương. Vì vậy, song song với chương trình cung cấp thông tin không thụ động thông qua hệ thống thông tin điện tử Internet (Trang Web điện tử) và hệ thống thông tin viễn thông thì việc ứng dụng các công nghệ, kỹ thuật hiện đại vào bảo quản lâu dài các tài liệu có giá trị (dần thay cho phương pháp bảo quản truyền thống) trở thành một nhiệm vụ cấp bách của công tác lưu trữ.

Hiện nay, một số Thư viện lớn trong nước như Thư viện Quốc gia, Thư viện tổng hợp Hồ Chí Minh, các Trung tâm lưu trữ đã ứng dụng việc lưu trữ tài liệu dưới dạng số hóa thông qua các phần mềm xử lý chuyên. Với hệ thống lưu trữ điện tử, tài liệu lưu trữ đã phát huy tối đa hiệu quả trong việc phục vụ vào lợi ích phát triển kinh tế, văn hóa và giáo dục truyền thống lịch sử cho thế hệ trẻ.

Nghệ An là tỉnh có diện tích lớn nhất Việt Nam thuộc vùng Bắc Trung Bộ, phía bắc giáp tỉnh Thanh Hóa, phía nam giáp tỉnh Hà Tĩnh, phía tây giáp Lào, phía đông giáp biển Đông. Nghệ An là mảnh đất có truyền thống hiếu học, có nhiều dòng họ, nhiều làng học nổi tiếng, là cái nôi sản sinh ra cho đất nước nhiều danh tướng, lương thần, nhiều nhà khoa học, nhà văn hóa có tầm cỡ quốc gia và quốc tế. Là nơi hội tụ và tích lũy các giá trị vật chất và tinh thần tạo nên một không gian văn hóa, một số di sản tài liệu quý hiếm, nhất là tài liệu Hán Nôm phong phú và đặc sắc của Việt Nam. Tuy nhiên, nhiều di sản tài liệu quý hiếm đó đang bị hủy hoại dần và có nguy cơ bị mất đi

theo thời gian. Khả năng tài liệu trên địa bàn tỉnh bị thất thoát và tiêu hủy trong thời kỳ chiến tranh, một số hư hỏng do thiên tai, lũ lụt và do khí hậu khắc nghiệt của miền Trung. Kho sách địa chí Thư viện Nghệ An có 10.000 bản trong đó có hàng trăm cuốn sách quý như Hán Nôm, sách báo xuất bản trong Kháng chiến chống Pháp... cũng khó bảo quản do thiếu phương tiện bảo quản hiện đại.

Vì vậy, trong bài viết này tôi mạnh dạn xin trao đổi một số vấn đề quan tâm hiện nay đối với công tác số hóa tài liệu trên địa bàn tỉnh Nghệ An vì mục đích bảo quản và phục vụ rộng rãi trong cộng đồng.

II. MỘT SỐ KHÁI NIỆM CƠ BẢN

Số hóa là gì? vì sao phải số hóa tài liệu?

Theo khái niệm của công nghệ thông tin thì số hóa là việc biến đổi các loại hình thông tin sang thông tin số (các bit thông tin dữ liệu). Các loại hình tài liệu (giấy, ảnh, phim...) sau khi qua công đoạn xử lý bằng các thiết bị chuyên ngành và phần mềm ứng dụng sẽ được số hóa thành các bit mang thông tin dữ liệu trên đường truyền Internet, tạo nên những cơ sở dữ liệu mở, dễ dàng tìm kiếm, trao đổi và chia sẻ kiếm thức một cách thuận tiện nhất.

Hiện nay, khối kiến thức khổng lồ của nhân loại hình thành trong hàng trăm năm nay, trong đó có cả tài liệu lưu trữ đang tồn tại dưới dạng tài liệu giấy như: Văn bản, sách, hình ảnh... trở thành một sự cản trở rất lớn cho người khai thác sử dụng bởi tính hữu dụng, khả năng tiếp cận xã hội rất hạn chế. Mặt khác, các tài liệu ở dạng này phải chịu tác động cơ học của con người, môi trường nên việc lưu giữ, kéo dài thời gian tuổi thọ của tài liệu đòi hỏi rất công phu, tốn kém. Và như đã nói ở trên, các dữ liệu được đưa lên hệ thống thông tin điện tử dưới dạng hàng hóa thông tin cũng phải tuân theo những quy luật cạnh tranh nhất định.

Tuy nhiên, chúng ta cũng dễ dàng nhận thấy, việc quản lý và cung cấp thông tin tài liệu lưu trữ điện tử chứa đựng những rủi ro như: Cơ sở dữ liệu bị xóa, thông tin bị chỉnh sửa... Chính vì vậy cần thiết kế một hệ thống lưu giữ tránh tình trạng xâm nhập trái phép, thực hiện chế độ quản lý tài liệu điện tử như là một bộ phận tổng thể các hồ sơ tài liệu, thông tin của cơ quan và cần có khuôn khổ chiến lược đối với tài liệu lưu trữ điện tử. Theo khái niệm của các chuyên gia, hệ thống lưu giữ tài liệu điện tử là một quy trình khép kín giúp các tài liệu được an toàn và được quản lý để tài liệu đó cùng với các thông tin, hoàn cảnh và cấu trúc của nó sẽ được giữ lại (Tính xác thực, độ tin cậy, tính an toàn, mối quan hệ với các đối tượng dữ liệu có liên quan, tính hữu dụng và khả năng tiếp cận). Tổ chức tiêu chuẩn quốc tế (ISO) đã chính thức ban hành tiêu chuẩn quốc tế về công tác văn thư ISO 15489, trong tiêu chuẩn này cũng đã đưa ra một chuẩn mực để các cơ quan, tổ chức có thể sử dụng nhằm đánh giá thực tiễn và hệ thống lưu trữ tài liệu điện tử.

Ải hư chúng ta đã biết, công tác số hóa tài liệu đã và đang đặt ra nhiệm vụ cấp bách trước những nguy cơ hư hỏng, xuống cấp trầm trọng do nhiều tác động trong đó có

môi trường. Vậy làm thế nào để bảo quản được vốn tài liệu này, đó là nhiệm vụ của chúng tôi và của tất cả các bạn nhất là những người làm công tác Thư viện.

Trước hết, để công việc này có hiệu quả, cần có sự quan tâm của lãnh đạo các địa phương, ban ngành và chú ý nâng cao ý thức của cán bộ, nhân viên về quản lý, bảo vệ tài liệu lưu trữ điện tử (Trong hệ thống quản lý chất lượng ISO ở cơ quan hành chính ở Hà Nội rất chú trọng đến vấn đề này) bằng cách xây dựng quy chế về quản lý tài liệu điện tử trong đơn vị, phân cấp một cách cụ thể trách nhiệm và quyền hạn xử lý, tiếp cận hồ sơ, tài liệu điện tử. Các tài liệu điện tử của mỗi cá nhân phải được lập hồ sơ công việc một cách rõ ràng dưới sự hỗ trợ của phần mềm ứng dụng (Ví dụ phần mềm hồ sơ công việc). Một yếu tố nữa quyết định chất lượng, hiệu quả của việc ứng dụng công nghệ thông tin vào bảo quản, khai thác sử dụng tài liệu lưu trữ điện tử là sự hỗ trợ đặc lực và phối hợp chặt chẽ của cán bộ tin học.

Để xây dựng tài liệu số phục vụ lâu dài, ổn định và hợp lý nhất thì ta phải sử dụng một giải pháp quản lý tài liệu số trên công giao tiếp điện tử mà sản phẩm của nó là các bộ sưu tập số, các cơ sở dữ liệu điện tử.

Vậy như thế nào là Bộ sưu tập số?

Bộ sưu tập số là một tập hợp có tổ chức nhiều tài liệu đã được số hóa dưới nhiều hình thức khác nhau (dạng text, hình ảnh, audio...) về một chủ đề nào đó.

Ấy vậy, một thư viện số có thể bao gồm nhiều bộ sưu tập theo các chủ đề khác nhau, có thể do tập thể hoặc cá nhân tự xây dựng hoặc trao đổi, mua bán; có thể nằm trong lưu trữ của đơn vị hoặc nằm ngoài đơn vị (thông qua một kênh cung cấp khác).

III. QUY TRÌNH THỰC HIỆN

Để công tác bảo quản, khai thác sử dụng tài liệu lưu trữ điện tử ít tốn kém, công việc đầu tiên chúng ta cần phải làm là hồ sơ, tài liệu hình thành trong xử lý công việc của từng cá nhân phải được phân loại và quản lý thống nhất trong hệ thống cơ sở dữ liệu. Tuyệt đối không tự ý xóa hoặc thay đổi thông tin của tài liệu, hàng năm tiến hành đánh giá, xử lý chất lượng và chống xâm nhập của tác nhân gây hại. Các hồ sơ, tài liệu điện tử đến hạn nộp lưu sẽ được chuyên giao đầy đủ cho cơ quan phụ trách lưu trữ phân loại, lập mã số điện tử bảo quản trong hệ thống lưu trữ điện tử; ấ vậy chúng ta sẽ giảm đi công đoạn tốn kém số hóa từ tài liệu giấy sang tài liệu điện tử.

1. Lựa chọn tài liệu

Đối với Thư viện, việc số hóa toàn bộ kho sách là điều khó thực hiện, đó là một thực tế không thể phủ nhận. Do đó, chúng ta cần phải đưa ra các tiêu chí làm căn cứ lựa chọn tài liệu cần thiết số hóa, cụ thể có những tiêu chí mà ta cần quan tâm:

**** Các loại tài liệu đặc biệt***

Là những loại tài liệu độc bản, tài liệu quý hiếm, tài liệu có thời gian xuất bản đã lâu như: Tài liệu Hán ấ, Địa chí, Tài liệu xuất bản trong kháng chiến chống Pháp

** Nội dung tài liệu*

Lựa chọn các chủ đề theo nội dung tài liệu như: Khoa học xã hội, Khoa học tự nhiên, nông nghiệp, Khoa học kỹ thuật, tài liệu có tần suất sử dụng cao

** Điều kiện bảo quản*

Tùy vào tình hình cụ thể của từng thư viện mà quyết định lựa chọn tài liệu đưa vào như ưu tiên các loại tài liệu trên giấy dễ rách nát

** Tình trạng bản quyền*

Vấn đề bản quyền thực sự đang là rào cản làm những người có tâm huyết với việc số hóa nản lòng. Điều đó xảy ra vì chúng ta chưa nắm vững những quy định cụ thể về việc bảo hộ quyền tác giả địa phương. Chúng ta có thể tham khảo Khoản a, đ Điều 25 Luật sở hữu trí tuệ năm 2005

** Nhóm người dùng*

Cũng tùy vào từng thư viện mà chúng ta xác định mức độ ưu tiên phục vụ như: Cán bộ lãnh đạo, cán bộ nghiên cứu, Cán bộ kỹ thuật, Sinh viên, Giáo viên và các đối tượng khác

2. Lựa chọn công nghệ

Việc lựa chọn công nghệ phải đáp ứng được các yêu cầu:

- Độ tin cậy
- Thuận tiện, dễ sử dụng
- Đáp ứng đầy đủ các yêu cầu về chuẩn nghiệp vụ Thư viện
- Phân quyền, bảo mật
- Trao đổi dữ liệu

Cơ sở hạ tầng: Hệ thống Internet, hệ thống máy chủ, cổng giao tiếp điện tử, phần mềm thư viện.

3. Số hóa nguồn tài nguyên

Đây là công đoạn đòi hỏi sự cẩn thận, tỉ mỉ qua nhiều giai đoạn và phải đầu tư nhiều công sức, kinh phí. Hiện nay thiết bị số hóa đã có những bước tiến rõ rệt, nếu trước đây để số hóa một cuốn sách dày 2000 trang thì phải mất gần 1 tuần để scan từng trang nhưng hiện nay cũng với cuốn sách đó chỉ mất vài giờ là cho ra một sản phẩm số đảm bảo chất lượng tốt, sắc nét, hình ảnh đẹp, giống 100% bản gốc và đặc biệt còn cho phép tự động tạo các siêu dữ liệu mô tả và siêu dữ liệu cấu trúc của tài liệu ở định dạng XML. Hiện nay ở Việt nam đã có các thiết bị số hóa tài liệu của công nghệ KIRTAS APT 1200, công nghệ này cùng với thiết bị BookScan APT 1200 có thể giúp các thư viện có thể số hóa nguồn tài liệu với số lượng lớn, giá cả hợp lý và đảm bảo chất lượng, thiết bị nhận dạng quang học OCR. Đặc biệt là công nghệ KIRTAS APT

1200 có một phần mềm biên tập BookScan Editor cho phép tự động biên tập, tạo siêu dữ liệu theo yêu cầu; BookScan APT 1200 không làm hư hỏng tài liệu gốc do không phải tháo gáy tài liệu đối với tài liệu có độ dày trang khi thực hiện Scan.

4. Biên mục và thiết lập các siêu dữ liệu liên kết

** Tạo siêu dữ liệu*

Có 3 dạng siêu dữ liệu: Siêu dữ liệu mô tả (mô tả các thông tin về tài liệu); Siêu dữ liệu cấu trúc (mô tả liên kết giữa các đối tượng thông tin như mục lục, chương, phần...) và Siêu dữ liệu quản trị (kích cỡ tập tin, định dạng tài liệu)

** Mô tả dữ liệu*

Có thể mô tả tài liệu theo một trong các chuẩn: MARC, Dublin Core..

Với sự phát triển mạnh mẽ của khoa học và công nghệ, cùng với xu thế hội nhập là những đòi hỏi và thách thức cho ngành Thông tin thư viện, nó đòi hỏi phải có những đổi mới hoạt động, bắt kịp những tiến bộ của thời đại, phục vụ đặc lực cho sự nghiệp công nghiệp hóa, hiện đại hóa đất nước.

IV. KẾT LUẬN

Để công tác bảo quản, khai thác sử dụng tài liệu lưu trữ điện tử ít tốn kém, công việc đầu tiên chúng ta cần phải làm là lập và xử lý hồ sơ, tài liệu hình thành trong xử lý công việc của từng cá nhân phải được phân loại và quản lý thống nhất trong hệ thống cơ sở dữ liệu. Tuyệt đối không tự ý xóa hoặc thay đổi thông tin của tài liệu, hàng năm tiến hành đánh giá, xử lý chất lượng và chống xâm nhập của tác nhân gây hại. Các hồ sơ, tài liệu điện tử đến hạn nộp lưu sẽ được chuyên giao đầy đủ cho cơ quan phụ trách lưu trữ phân loại, lập mã số điện tử bảo quản trong hệ thống lưu trữ điện tử; ắt hẳn vậy chúng ta sẽ giảm đi công đoạn tốn kém số hóa từ tài liệu giấy sang tài liệu điện tử.

Xây dựng hệ thống bảo quản, khai thác sử dụng tài liệu điện tử song song với bảo quản tài liệu giấy hiện nay trở thành vấn đề rất được quan tâm của các quốc gia tiên tiến. Đối với chúng ta, để công tác này đi vào nề nếp, cần có sự chỉ đạo thống nhất, sự nỗ lực của các địa phương, ban ngành và ý thức trách nhiệm của từng cán bộ công chức đối với tài liệu điện tử nhằm từng bước đưa công tác này theo hướng hiện đại./.