

Một số kinh nghiệm số hóa tài liệu tại Trung tâm Lưu trữ quốc gia II

Vũ Văn Tâm

Trưởng phòng Tin học và Công cụ tra cứu Trung tâm Lưu trữ quốc gia II

1. VÀI NÉT VỀ SỐ HÓA TÀI LIỆU LƯU TRỮ TẠI TRUNG TÂM LƯU TRỮ QUỐC GIA II

Trung tâm Lưu trữ quốc gia II tại thành phố Hồ Chí Minh là đơn vị sự nghiệp thuộc Cục Văn thư và Lưu trữ nhà nước – nơi đang bảo quản trên 14.000 mét giá tài liệu có thời gian từ năm 1862 đến năm 2000, gồm các loại hình tài liệu: tài liệu hành chính, tài liệu khoa học kỹ thuật, tài liệu nghe nhìn và sưu tập tài liệu cá nhân.

Trong những năm qua được sự quan tâm của các cấp lãnh đạo, Trung tâm II thực hiện bảo quản an toàn, tổ chức sử dụng có hiệu quả tài liệu lưu trữ, phục vụ nhu cầu khai thác của độc giả và góp phần vào công cuộc xây dựng đất nước. Tuy nhiên, cùng với các yếu tố khách quan, việc khai thác trực tiếp trên tài liệu trở thành nguyên nhân góp phần làm tài liệu lưu trữ ngày càng xuống cấp. Đồng thời, việc khai thác trực tiếp trên tài liệu không thể đáp ứng kịp thời các yêu cầu phát triển của xã hội với sự phát triển mạnh mẽ, không ngừng của CNTT. Do đó, yêu cầu số hóa tài liệu nhằm đáp ứng hai mục tiêu trên trở thành yêu cầu cấp bách không chỉ ở Việt Nam mà còn mang tính toàn cầu.

Với nhận thức đó, Cục Văn thư và Lưu trữ Nhà nước sớm có những chỉ đạo thực hiện số hóa tài liệu tại các Trung tâm Lưu trữ quốc gia. Riêng tại Trung tâm Lưu trữ quốc gia II, thực hiện trách nhiệm được giao, năm 1998, Trung tâm bắt đầu thực hiện dự án số hóa tài liệu với dự án số hóa khối tài liệu Mộc bản triều Nguyễn. Trong điều kiện thiếu thốn về nhân vật lực, vừa làm vừa học hỏi, bằng các loại máy quét tĩnh (Flatbed scanner) và sử dụng hai dạng phần mềm: Phần mềm quét ảnh đi kèm với máy quét và phần mềm xử lý hình ảnh Adobe Photoshop, Trung tâm đã tiến hành quét được 55.000 trang tài liệu, lưu trong 184 đĩa CD ROM. Tiến hành xây dựng phần mềm quản lý, đưa dữ liệu số hóa tài liệu Mộc bản triều Nguyễn vào phục vụ kịp thời nhu cầu khai thác của độc giả.

Từ những thành công bước đầu, năm 2001, Trung tâm II tiến hành số hóa khối tài liệu Ghi âm – là những băng ghi âm về những sự kiện, hoạt động của chính quyền Nguyễn Văn Thiệu, được Trung tâm II thu gom về từ năm 1977. Đây là những cuộn băng do Sở Tin báo – Nha Báo chí – Phủ Tổng thống thực hiện và lưu giữ, phần lớn không được in sao hoặc in thành văn bản (tức chỉ có 01 bản duy nhất) với số lượng 600 cuộn. Tất cả tài liệu ghi âm ở trong tình trạng xếp đống, chưa được phân loại và bị xuống cấp trầm trọng (băng bị nấm mốc, quấn xoắn, bết dính và đã có hội chứng dấm chua,... dẫn đến chất lượng âm thanh rất kém), đứng trước nguy cơ bị hư hỏng hoàn toàn. Việc cứu nguy khối

tài liệu Ghi âm trở thành thách thức lớn đối với đội ngũ CBVC Trung tâm II. Giải pháp số hóa nhằm cứu nguy khối tài liệu này được đề ra. Song những kinh nghiệm số hóa tài liệu Mộc bản triều Nguyễn (tài liệu giấy) trước đây không đủ đáp ứng yêu cầu số hóa cho loại hình tài liệu mới. Những câu hỏi về giải pháp được đặt ra: Phương pháp xử lý âm thanh? Lựa chọn định dạng file? Lựa chọn vật liệu ghi số: ghi lên ổ cứng, ghi lên băng từ số hay ghi lên đĩa quang là những bài toán về tính tiện ích, giá thành và tuổi thọ. Đặc biệt là giải pháp xử lý âm thanh gặp hai trở ngại lớn: một là, vấn đề thiết bị đọc cho loại băng đĩa có từ trước năm 1975 — loại băng cối chỉ đọc được bằng máy AKAI, mà việc tìm kiếm trên thị trường là vô cùng khó khăn. Vấn đề đồng bộ hóa máy AKAI với máy vi tính sau đó lại nảy sinh; hai là, việc xử lý tạp âm, khi hầu hết các cuộn băng được ghi ngoài trời, hoặc trong các cuộc họp báo với rất nhiều loại ngôn ngữ, tạp âm khác nhau,....

Nhưng bằng sự nỗ lực của đội ngũ CBVC, với phương châm đúc rút kinh nghiệm từ thực tiễn, sau nhiều lần thử nghiệm đến cuối năm 2004, Trung tâm II thực hiện thành công việc số hóa khối tài liệu Ghi âm. Với kết quả đã số hóa và xử lý kỹ thuật âm thanh được 429 giờ sản phẩm lưu trữ trong 125 đĩa CD-ROM. Đồng thời, Trung tâm cũng xây dựng thành công phần mềm tích hợp thông tin cấp 1 và cấp 2 (có thể tra cứu đồng thời bản ghi âm và bản text), đáp ứng nhu cầu khai thác của độc giả.

Năm 2008, Trung tâm II đẩy mạnh hơn nữa việc số hóa tài liệu bằng việc thực hiện dự án số hóa tài liệu hành chính và chuyển ảnh số sang vi phim. Đến cuối năm 2008, đúc kết kinh nghiệm từ nhiều lần thực nghiệm, Trung tâm II đã xây dựng thành công quy trình số hóa tài liệu hành chính và chuyển sang vi phim với yêu cầu tạo ra hai dạng dữ liệu (dạng tiff để chuyển sang vi phim và dạng jpeg để phục vụ khai thác), đòi hỏi trang thiết bị, kỹ thuật phù hợp.

Quy trình số hóa tài liệu tại Trung tâm Lưu trữ quốc gia II

Sơ lược một số bước trong quy trình số hóa như sau:

1. Chuẩn bị tài liệu số hóa:
 - Thống kê tài liệu;
 - Tháo bỏ toàn bộ ghim, kẹp;
 - Ủi phẳng tài liệu;
 - Lập danh mục tài liệu (biên mục từng văn bản trong hồ sơ);
 - Kiểm tra, đối chiếu tài liệu với danh mục, xác định các tiêu chụp đặc biệt, xác định số lượng hồ sơ trong 1 cuộn phim và lập “Mục lục” hồ sơ cho cuộn phim;
 - Tạo dải ảnh để số hóa.

2. Số hóa tài liệu:

- Trước khi thực hiện quét ảnh, phải điều chỉnh độ tương phản sáng tối, độ bóng của chữ,... tùy theo tình trạng cụ thể của mỗi trang tài liệu và máy quét sử dụng;
- Dung lượng mỗi ảnh không vượt quá 1,2MB;
- Đối với tài liệu quét có kích thước lớn hơn khổ A4 (vượt quá vùng đặt tài liệu để quét của máy) thì phải chia tài liệu thành hai phần và quét hai lần theo thứ tự từ trên xuống dưới, phần liên kế gối lên nhau khoảng 20 mm. Số thứ tự của ảnh được xác định theo thứ tự quét ảnh;
- Tiêu chuẩn hình ảnh được sử dụng tại Trung tâm Lưu trữ quốc gia II như sau:

+ Đối với bản sao phục vụ khai thác sử dụng:

- Áp dụng tiêu chuẩn: Joint Photographic Expert Group (.jpg);
- Ảnh màu;
- Độ phân giải tối thiểu: 100 dpi (tùy thuộc vào tình trạng tài liệu);
- Tỷ lệ quét: 25%-100%.

+ Đối với bản sao phục vụ lập bản sao bảo hiểm (ghi sang microfilm):

- Áp dụng tiêu chuẩn: Tag Image File (.tif); kỹ hiệu: TIFF;
- Ảnh đen trắng;
- Độ phân giải tối thiểu: 300 dpi;
- Tỷ lệ quét: 100%.

3. Đặt tên FILE:

Tên file ảnh được đặt thống nhất theo quy định tại Quyết định số 324/QĐ-VTLTNN ngày 26 tháng 8 năm 2005 của Cục Văn thư và Lưu trữ Nhà nước ban hành Quy định về chuẩn thông tin đầu vào và biên mục văn bản, tài liệu hành chính tại các Trung tâm Lưu trữ quốc gia gồm: Mã cơ quan lưu trữ_phông số_mục lục số_hồ sơ số_tờ số_mặt số.

4. Kiểm tra chất lượng ảnh:

- Kiểm tra chất lượng toàn bộ hình ảnh đã quét, đặc biệt là những hình ảnh cần chuyển sang microfilm

Chú ý:

- Bảo đảm các hình ảnh sắc nét và rõ ràng;
- Nên ảnh phải sạch, không bị lốm đốm, “tạp nhiễu”;

- Các hình ảnh được quét xếp theo cùng hướng;
- Hình ảnh phải được quét vuông góc, không bị cong vênh;
- Các hình ảnh được sắp xếp theo đúng trật tự của tài liệu gốc.

– Quét lại những ảnh không đạt yêu cầu, những ảnh quét lại phải được đặt tên với đúng tên file ảnh mà nó thay thế.

5. Sao lưu dữ liệu

6. Tạo siêu dữ liệu kết nối:

Dữ liệu số hóa sẽ được liên kết đến hồ sơ tương ứng trên cơ sở dữ liệu.

Từ năm 2009, trên cơ sở xây dựng được quy trình số hóa tài liệu, Trung tâm II triển khai thực hiện dự án số hóa tài liệu hành chính và chuyển sang vi phim với tất cả các phòng tài liệu đã được chính lý hoàn chính.

2. MỘT SỐ YÊU CẦU VỀ SỐ HÓA TÀI LIỆU

Thành quả của việc số hóa tài liệu là vô cùng to lớn, góp phần quan trọng vào bảo vệ, bảo quản an toàn bản gốc tài liệu, tạo ra phương thức mới về khai thác, sử dụng tài liệu lưu trữ. Song, số hóa tài liệu cũng tiềm ẩn nguy cơ về sự sai lạc thông tin so với bản gốc và đặc biệt có nguy cơ cao về việc để mất cơ sở dữ liệu. Vì vậy, nó có những đòi hỏi cao về quá trình bảo quản, sử dụng kho dữ liệu số và tính chính xác trong quá trình thực hiện.

• ĐỐI VỚI BẢO QUẢN VÀ SỬ DỤNG KHO DỮ LIỆU SỐ HÓA

Tài liệu số không chỉ có ý nghĩa bảo hiểm các loại hình tài liệu khác, mà trên phương diện nhất định, đây cũng là một loại hình tài liệu – dạng tài liệu đặc thù, cần có phương pháp bảo quản và sử dụng riêng.

Về bảo quản tài liệu số đặt ra một số vấn đề:

Bảo quản tài liệu số có nhiều vấn đề hơn để đối phó với các tiến bộ về công nghệ và tổ chức đang được thay đổi một cách nhanh chóng. Thách thức ban đầu đối với công tác bảo quản dữ liệu số là vấn đề thiết bị lưu trữ các tài liệu số rất dễ bị tổn hại và mất mát so với tài liệu analog (như tài liệu giấy, ...).

Do đó cần đặt ra chiến lược bảo quản tài liệu lâu dài bằng kỹ thuật số phải được đầu tư, duy trì và phát triển như bảo quản tài liệu analog khác.

Tất cả các tài liệu lưu trữ số sẽ được bảo quản theo mục đích sử dụng cao hay thấp. Đối với bản sao số gốc sẽ được bảo quản nghiêm ngặt bởi nhiệt độ và độ ẩm theo tiêu chuẩn và có thể được sao lưu trên nhiều thiết bị lưu trữ khác nhau như đĩa quang, băng từ, ổ cứng. Đối với bản sao sử dụng có độ phân giải thấp được lưu trữ tại server và những bản lưu trữ dự phòng.

Song song với bảo quản, Tổ chức sử dụng tài nguyên số cũng cần đảm bảo duy trì thông tin trong kho lưu trữ; đảm bảo tính chính xác, trung thực, khách quan; đảm bảo tính khoa học, thuận tiện cho khai thác và sử dụng; đảm bảo tính kịp thời, đầy đủ, có hệ thống và hướng tới các mục tiêu:

- Tăng cường khả năng tiếp cận bằng việc sử dụng bản sao kỹ thuật số;
- Mở rộng đối tượng phục vụ; ...

Để đạt được các mục tiêu trên, yêu cầu đặt ra là cần có các phần mềm tiện dụng, đáp ứng được cả yêu cầu quản lý và nhu cầu sử dụng.

- **ĐỐI VỚI QUÁ TRÌNH THỰC HIỆN:**
- Phải xác định rõ mục tiêu số hóa tài liệu (để phục vụ thay vì sử dụng bản gốc, để lập bản sao bảo hiểm hay cả hai mục đích trên) từ đó để lựa chọn thiết bị và xây dựng các tiêu chuẩn số hóa cho phù hợp;
- Phải chuẩn bị tốt đội ngũ cán bộ vì đây là khâu then chốt quyết định sự thành bại;
- Phải chuẩn bị tốt cơ sở vật chất (máy móc, thiết bị, vật tư, kinh phí, ...) để triển khai việc số hóa tài liệu;
- Phải chuẩn hóa các quy trình nghiệp vụ (quy trình số hóa, xác định các chuẩn thông tin đầu vào, đầu ra) để làm cơ sở cho việc thiết kế phần mềm ứng dụng cho phù hợp;
- Dữ liệu ảnh số hóa phải được ghi trên các vật mang tin có chất lượng cao. Cần được chia sẻ tạo điều kiện để phát huy giá trị của tài liệu lưu trữ;
- Cần tiến hành nghiên cứu và đưa ra quy trình bảo quản tài liệu số;
- Cần thường xuyên nâng cao trình độ của cán bộ để có thể theo kịp với các thay đổi của công nghệ;
- Cần thường xuyên theo dõi sự phát triển của công nghệ để có thể đối phó được sự lỗi thời của các định dạng kỹ thuật số, phần cứng và phần mềm;
- Cần đặt ra thời gian thích hợp để kiểm tra, nâng cấp và có quy định cụ thể về thời gian sao lưu dữ liệu sang các phương tiện khác.

Nhìn chung, được sự quan tâm đầu tư của các cấp lãnh đạo, việc số hóa tài liệu tại Trung tâm II đã đạt được những thành quả bước đầu khả quan. Các tài liệu được số hóa đã đáp ứng được yêu cầu đặt ra, góp phần bảo vệ nguồn tài liệu gốc khỏi bị hủy hoại do tác động của hóa lý trong quá trình sử dụng lâu dài; cũng như phục vụ nhu cầu, mong muốn được tiếp cận tài liệu lưu trữ một cách nhanh chóng và thuận tiện của độc giả. Song việc số hóa tài liệu cũng còn tồn tại một số khía cạnh cần cải tiến về quy trình, kỹ thuật, đặc biệt là trong sao lưu, bảo quản, bảo mật kho dữ liệu số. Do đó, việc đầu tư, tìm kiếm giải pháp công nghệ vẫn luôn là vấn đề then chốt.