

# **SỐ HÓA TÀI LIỆU TRONG XU THẾ PHÁT TRIỂN THƯ VIỆN SỐ TẠI THƯ VIỆN KHOA HỌC XÃ HỘI**

*PHÙNG THỊ BÌNH<sup>(\*)</sup>*

1. Thư viện KHXH là thư viện đầu ngành của Việt Nam về KHXH. Với nguồn thông tin phong phú về các ngành KHXH, Thư viện KHXH là nơi trau dồi tri thức của nhiều thế hệ độc giả.

Thư viện KHXH được kế thừa nguồn tài liệu rất quý từ Viện Viễn Đông Bác cổ Pháp (EFEO) để lại, bao gồm: kho Hương ước, Thần tích Thần sắc, tài liệu Latinh cũ, sách Trung Quốc cổ, sách Hán Nôm, sách Nhật Bản cổ, kho ảnh, kho phim, bản đồ, sắc phong, đĩa hát và kho tranh ảnh cổ. Các kho sách của EFEO để lại đã được các nhà nghiên cứu khẳng định rất có giá trị trong nghiên cứu KHXH. Các kho sách hình thành từ khi thành lập Thư viện Khoa học Trung ương (06/02/1960) đến nay cũng rất đa dạng và phong phú phục vụ nhu cầu nghiên cứu khoa học của nhiều thế hệ các nhà khoa học trong những năm qua. Đến nay, thư viện đã có trên 1.000.000 bản sách và báo chí ấn bản, gồm nhiều ngôn ngữ khác nhau. Do đó, yêu cầu số hóa nhằm phổ cập đến đông đảo người đọc và để bảo quản là rất cần thiết.

Tồn tại trên 100 năm, Thư viện đang từng ngày phát triển bất kịp xu thế của các thư viện trong nước và trên thế giới. Trong định hướng phát triển, Thư viện KHXH đặt mục tiêu xây dựng thư viện số về KHXH vào năm 2020. Do vậy, Thư viện đang tích cực chuẩn bị về mọi mặt để hoàn thành mục tiêu đặt ra, trong đó xây dựng nguồn tài liệu số là mục tiêu được quan tâm.

Nguồn tài liệu số là nguồn tài nguyên quan trọng lưu thông của thư viện số, đáp ứng mọi nhu cầu của độc giả trên khắp thế giới. Với mỗi thư viện, nguồn tài liệu được coi là linh hồn của chính thư viện đó, phản ánh đặc thù, giá trị và tầm quan trọng của thư viện. Như vậy, việc số hóa có chọn lọc nguồn tài liệu của thư viện là lựa chọn chính bên cạnh khả năng bổ sung tài liệu số từ các nguồn khác. Mặt khác, số hóa còn là giải pháp bảo tồn, bảo quản tài liệu được đánh giá cao, đặc biệt là với những tài liệu cổ, độc bản, tài liệu di sản, tài liệu có thời gian xuất bản lâu năm đang

---

<sup>(\*)</sup> ThS., Viện Thông tin KHXH.

trong tình trạng bị xuống cấp. Nguồn tài liệu số giúp nâng cao khả năng đáp ứng nhu cầu thông tin cho độc giả, trong đó độc giả của Thư viện KHXH chủ yếu là các nhà nghiên cứu. Loại hình tài liệu số có ưu điểm là: kiểm soát tài nguyên thông tin; bảo vệ an toàn và lâu dài các tài liệu gốc; nâng cao năng lực khai thác thông tin của người dùng tin; dễ dàng tạo lập các loại sản phẩm và dịch vụ thông tin mới; thúc đẩy mở rộng việc chia sẻ thông tin. Trong bối cảnh Thư viện đang chuẩn bị cơ sở hạ tầng và điều kiện để xây dựng thư viện số, một thư viện hiện đại thông minh thì tài liệu số chiếm vị trí đặc biệt quan trọng.

**2. Tại Thư viện KHXH, nguồn tài liệu số đã tăng lên nhanh chóng trong những năm qua. Công tác số hóa tài liệu - một trong những cách quan trọng nhất để phát triển tài liệu số - được chú trọng. Thư viện đã có bộ phận số hóa chuyên nghiệp tại phòng Xây dựng Cơ sở dữ liệu - Thư mục. Thư viện KHXH đã đưa nghiệp vụ số hóa vào hoạt động thường xuyên. Các công tác số hóa đang được nghiên cứu tích cực nhằm đưa tài liệu số đạt tới chuẩn chung của thế giới.**

Trong những năm qua, công tác số hóa tài liệu của Thư viện đã được các cấp lãnh đạo quan tâm, Thư viện được đầu tư các trang thiết bị số hóa khá đa dạng và hiện đại, phù hợp với phần lớn kích thước tài liệu trong Thư viện, bao gồm:

- 2 máy ảnh kỹ thuật số Sony: có khả năng chụp ảnh tài liệu có kích thước lớn đến 1m. Máy ảnh cho kết quả là ảnh RAW, JPEG có độ phân giải trên 300ppi, dung lượng trên 14MB, đạt tiêu chuẩn sử dụng để lưu trữ lâu dài.

- 1 máy scan A4, 2 máy scan A3, 1 máy Scan Robot: Các thiết bị này có thể số hóa tài liệu với nhiều kích thước khác nhau. Các thiết bị scan này đều cho đa dạng các dạng ảnh như TIFF, JPEG, PDF,... với độ phân giải lên tới 600ppi.

- 1 máy scannap SV600 của hãng Fujitsu Nhật Bản.

Với sự đa dạng về thiết bị trên, bộ phận số hóa có điều kiện thực hiện số hóa tài liệu trên nhiều dạng kích thước tài liệu khác nhau. Có thể số hóa tài liệu có kích thước đến A<sub>0</sub>. Năm 2014, Thư viện KHXH đã số hóa được bản đồ khổ lớn (với kích thước lên tới khoảng 1,6m x 2m) với chất lượng cao, đáp ứng yêu cầu bảo tồn và xuất bản lại.

Phương pháp số hóa tài liệu thực chất là lựa chọn các thiết bị số hóa phù hợp nhất đối với các loại tài liệu khác nhau nhằm đạt chất lượng file tốt nhất và giảm thiểu những tác động làm hư hại thêm cho tài liệu. Để chất lượng tài liệu số đạt kết quả cao, đồng thời đảm bảo được sách gốc khi số hóa không bị tác động làm xuống cấp thêm, Thư viện KHXH luôn lựa chọn các phương pháp số hóa hợp lý nhất. Có hai hình thức số hóa tài liệu đang được áp dụng tại Thư viện (theo quan điểm phân chia của chúng tôi):

- Số hóa thủ công: sử dụng các thiết bị như máy ảnh, máy scan không có phần mềm xử lý chuyên biệt đi kèm, ảnh sau khi chụp/quét được xử lý trên một phần mềm đơn lẻ.

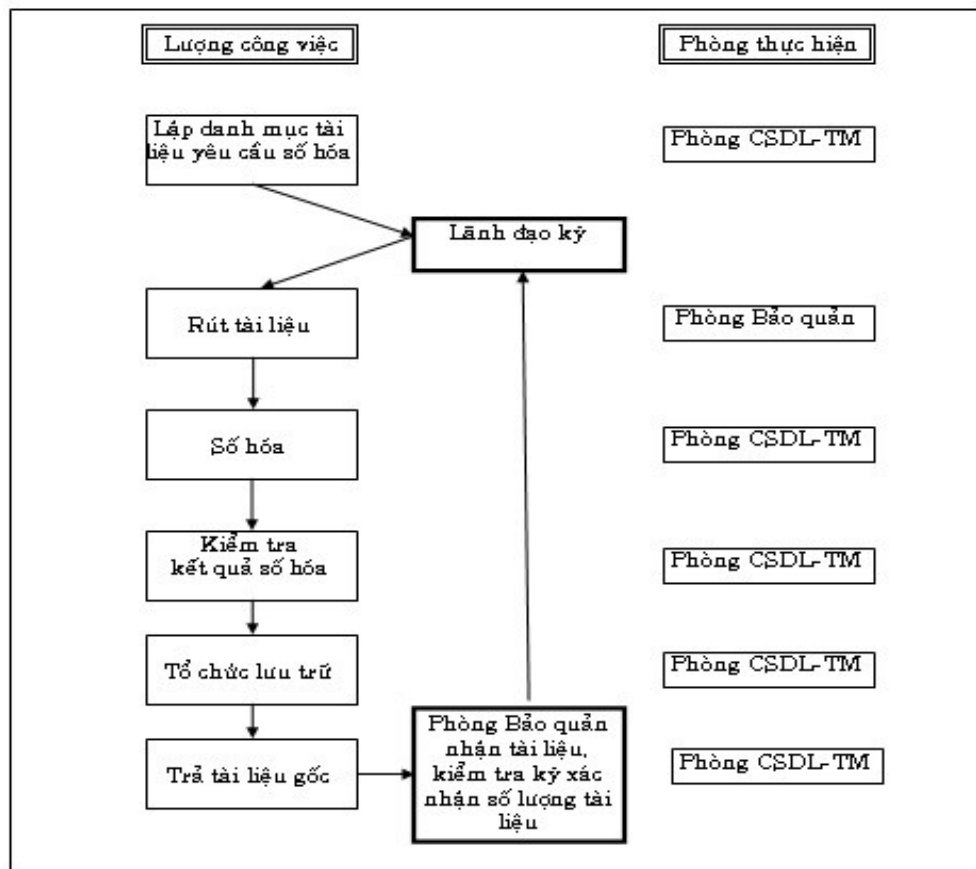
- Số hóa tự động: sử dụng thiết bị quét tự động như scanrobot, ảnh được xử lý bằng chính phần mềm tích hợp của hệ thống.

Dù sử dụng hình thức số hóa nào thì các sản phẩm số hóa cũng phải đạt các tiêu chuẩn kỹ thuật về tài liệu số hóa, đảm bảo các quy tắc định danh, đảm bảo sau khi số hóa các bản tài liệu gốc vẫn an toàn, không rách nát, không mất trang, không mất/lạc tài liệu.

**3.** Về nguyên tắc, số hóa tài liệu phải tuân thủ theo một quy trình. Quy trình số hóa là một chuỗi các hoạt động liên tiếp được thực hiện từ khâu đầu tiên đến khi cho ra sản phẩm cuối cùng theo yêu cầu. Trước đây, khi số hóa tài liệu theo các đề tài, dự án, quy trình chỉ được chú ý từ khi quét tài liệu đến khi ra sản phẩm cuối cùng, chưa đề cập đến các khâu trước khi quét tài liệu và sau khi hoàn thành sản phẩm. Bên cạnh đó,

việc số hóa một khối lượng tài liệu lớn phải chạy theo thời gian dẫn đến không kiểm soát được quy trình thực hiện của mỗi cán bộ cũng là nguyên nhân dẫn đến làm thay đổi các thông số kỹ thuật của tài liệu số.

Một quy trình thành văn, chính thức quy định các bước thực hiện số hóa tài liệu là rất cần thiết. Trong đó, quy trình quy định chặt chẽ ngay từ bước đầu tiên là lập danh mục những tài liệu sẽ được số hóa, đề xuất lãnh đạo duyệt, sau đó chuyển tới bộ phận rút tài liệu từ kho bảo quản. Như vậy, trước khi thực hiện các thao tác số hóa cụ thể, quy trình đã phải qua nhiều bước, điều này là bắt buộc. Vì quy trình thực hiện càng chặt chẽ bao nhiêu thì độ an toàn của



tài liệu gốc càng được đảm bảo. Hơn nữa, công tác số hóa tại Thư viện KHXH có liên quan đến các phòng nghiệp vụ khác trong toàn bộ dây chuyền thư viện, nên những chi tiết được quy định chặt chẽ là cơ sở để các cán bộ thực hiện số hóa và các phòng liên quan thực hiện tốt nhiệm vụ được giao. Sơ đồ mô tả quy trình số hóa tài liệu tại Thư viện KHXH thể hiện ở trang 79.

Thông thường, các thao tác trong quy trình số hóa được hướng dẫn trực tiếp cho cán bộ thực hiện. Quy trình chưa được biên soạn thành văn bản chính thức của Thư viện nên công tác số hóa cũng gặp một số khó khăn. Thư viện thường triển khai số hóa tài liệu với nhiều cán bộ cùng tham gia, đa số là các cán bộ làm chuyên môn khác nên có thể dẫn đến thực hiện sai quy trình, bẻ gãy quy trình hoặc biến đổi quy trình, đặc biệt là những thao tác xử lý hình ảnh. Những bất cập này là nguyên nhân dẫn đến chất lượng sản phẩm số hóa không đồng đều và một số sản phẩm chưa đạt chuẩn.

Thời điểm hiện tại, tài liệu số tại Thư viện đã bắt đầu được đưa vào sử dụng, nhưng mới chỉ trong phạm vi hẹp, chủ yếu trên mạng LAN, chưa truy cập được qua mạng Internet. Dữ liệu số được cung cấp cho độc giả khi có yêu cầu cung cấp file hoặc in ấn. Danh mục nguồn tài liệu số tại Thư viện KHXH (tính đến hết năm 2014) thể hiện ở bảng trang 81.

**4. Sử dụng tài liệu số trong Thư viện** là hình thức đáp ứng nhu cầu tin của độc giả không phải bằng tài liệu dạng vật lý mà bằng tài liệu được số hóa lưu trữ dưới dạng các file, các tệp tin máy tính. Có thể phục vụ độc giả trực

tiếp đọc trên máy tính của Thư viện hoặc từ xa thông qua mạng Internet. Độc giả có thể yêu cầu được phục vụ bản file hoặc in lại tùy mục đích sử dụng.

Đưa tài liệu số vào sử dụng cũng là nhằm hạn chế sử dụng bản gốc để tránh gây hư hại cho tài liệu, bảo quản tài liệu gốc được lâu dài. Hiện tại, một phần của kho tài liệu số của Thư viện KHXH đã được đưa vào sử dụng đáp ứng nhu cầu độc giả.

Tài liệu sau khi số hóa cần được lưu trữ và bảo quản theo các yêu cầu về an ninh dữ liệu. Để thực hiện được các giải pháp an ninh dữ liệu, trước hết Thư viện cần trang bị được một hệ thống lưu trữ chuyên dụng, có độ ổn định cao. Đối với một thư viện có kho tài nguyên số lớn thì việc lựa chọn giải pháp lưu trữ phù hợp là vô cùng quan trọng. Trong trường hợp của Thư viện KHXH, Thư viện đang định hướng phát triển thư viện số những năm tiếp theo, thì một hệ thống lưu trữ và quản trị được toàn bộ nguồn tài nguyên ngày càng tăng của Thư viện càng cần phải được lựa chọn chính xác. Hiện tại trên thị trường có 3 hệ thống lưu trữ: *DAS* (Direct attached storage), *NAS* (Network Attached Storage), và *SAN* (Storage Area Network). Ứng dụng cho lưu trữ và quản trị dữ liệu, lựa chọn phù hợp nhất cho Thư viện là giải pháp SAN, bởi phân tích các ưu điểm của SAN cho thấy hệ thống này hoàn toàn đáp ứng được các yêu cầu của một hệ thống lưu trữ dữ liệu, gồm:

- Hệ thống lưu trữ có khả năng tích hợp được các loại công nghệ lưu trữ, thiết bị lưu trữ và có thể hỗ trợ được nhiều loại hình lưu trữ.

<b>Danh mục nguồn tài liệu số tại Thư viện KHXH</b>				
<b>TT</b>	<b>Tên bộ sưu tập</b>	<b>Dạng số hóa</b>	<b>Số lượng</b>	<b>Ghi chú</b>
1	Hương ước	Số hóa toàn văn. File ảnh: Tiff, jpeg, PDF	5.461 cuốn	Số hóa trọn bộ sưu tập
2	Hương ước Nôm	Số hóa toàn văn. File ảnh: Tiff, jpeg, PDF	1.226 cuốn	Số hóa trọn bộ sưu tập
3	Thần sắc Hán Nôm	Số hóa toàn văn. File ảnh: Tiff, jpeg, PDF	3.710 cuốn	Số hóa trọn bộ sưu tập
4	Sắc phong	Chụp ảnh, file ảnh: jpeg	533 bản	Số hóa trọn bộ sưu tập
5	Ảnh	Quét file ảnh: Tiff, jpeg.	57.860 ảnh	Số hóa trọn bộ sưu tập
6	Tạp chí <i>L'éveil économique de l'Indochine</i>	File PDF	625	215 số chưa số hóa
7	Sách Hán Nôm	Chụp ảnh file RAW, jpeg, PDF	243 cuốn toàn văn 1.496 cuốn 3 trang	243 cuốn toàn văn, còn lại 3-15 trang
8	Sách Trung Quốc cổ	Chụp/quét ảnh file: jpeg, PDF	42.143 cuốn	3-5 trang
9	Sách Nhật Bản cổ	Chụp/quét ảnh file: jpeg, PDF	9.693 cuốn	3-5 trang
10	Sách Pháp	Quét ảnh file: PDF	300 cuốn	Toàn văn
11	Bản đồ	Quét ảnh file: Tiff, jpeg	49 bản đồ	Chủ đề bản đồ Điện Biên Phủ và Hoàng Sa, Trường Sa
12	Sách mới	fulltext	320 cuốn 2.800 số hóa 10 -15 trang	
13	Kết quả các đề tài nghiên cứu khoa học	Fulltext	Trên 200 đề tài	
14	Tài liệu phục vụ nghiên cứu	HTML, PDF	Khoảng 2.700 số	Từ 1990-2000 file ảnh 2000 -2013 fulltext
15	Công báo	Chụp/quét file: JPEG, PDF	110 tập	
16	OCTO	Chụp/quét file: TIFF, JPEG, PDF	42 cuốn	Sách in trước năm 1945 bằng tiếng Việt
17	Các tài liệu khác	File PDF	27 loại	Tài liệu biếu tặng

- Xem xét và nắm bắt xu hướng phát triển công nghệ của các nhà sản xuất lớn nhằm cung cấp dịch vụ lưu trữ tốt nhất cho khách hàng.

- Hệ thống đáp ứng yêu cầu lưu trữ hiện tại và trong tương lai.

- Hệ thống có độ sẵn sàng và khả năng thực thi ứng dụng cao.

- Có thể mở rộng phát triển một cách liên tục mà vẫn đảm bảo đầu tư hợp lý.

- Có hệ thống dịch vụ chuyên nghiệp và giải pháp hoàn thiện.

- Dễ triển khai và quản trị: Phần mềm quản lý trung tâm sẽ làm đơn giản hoá công tác quản trị và giảm được các chi phí vận hành (<http://www.quangngai.gov.vn/>).

SAN chính là hệ thống lưu trữ hiện đại đáp ứng được các vấn đề đặt ra của công tác lưu trữ bảo quản tài liệu kỹ thuật số, hỗ trợ đắc lực cho sự vận hành hệ thống dữ liệu thư viện số tại Thư viện KHXH. Đây có thể sẽ là sự lựa chọn của hầu hết các thư viện trong thời gian tới.

Trong bảo quản, việc kéo dài tuổi thọ của tài liệu (vật mang tin) luôn là vấn đề được quan tâm. Tuy nhiên, đối với tài liệu kỹ thuật số, tuổi thọ của các tài liệu này lâu bền hơn so với tài liệu vật lý. Chính vì thế, việc tồn tại của các tài liệu này phụ thuộc nhiều vào tuổi thọ của hệ thống truy cập. Nói cách khác, đó chính là khả năng khai thác tài liệu qua các hệ thống. Do không thể chắc chắn về thời hạn hoặc khả năng bảo trì, hỗ trợ của nhà cung

cấp đối với một hệ thống, nên các thư viện cần sẵn sàng chuyển đổi những dữ liệu ảnh có giá trị, những dữ liệu chỉ mục và phần mềm sang những thế hệ công nghệ tiếp theo.

Trên thực tế, những trung tâm lưu trữ lớn như Trung tâm Lưu trữ Quốc gia vẫn sử dụng công nghệ lưu trữ trên Microfilm của thế kỷ trước. Nếu chỉ lưu trữ tài liệu dưới góc độ bảo tồn và phục vụ độc giả theo phương thức truyền thống thì lưu trữ trên microfilm là giải pháp tối ưu. Cho đến nay, microfilm vẫn được cho là có tuổi thọ khá cao (khoảng 500 năm), trong khi các hệ thống lưu trữ và công nghệ hiện đại có tuổi thọ khá thấp. Đặc biệt, các thiết bị công nghệ thông tin lỗi thời rất nhanh, đó cũng là một thách thức lớn cho các thư viện trên thế giới nói chung và Thư viện KHXH nói riêng.

Nhìn vào thực tế của Thư viện KHXH, nên chăng bên cạnh việc tiến tới trang bị những hệ thống lưu trữ, bảo quản tài liệu hiện đại thì cần cần nhắc một khả năng ghi tài liệu số trên microfilm để lưu trữ bảo quản. Đó là một giải pháp an toàn cho những dữ liệu quý, độc bản đã số hóa toàn văn: Thần sắc Hán Nôm, Hương ước Nôm, Hương ước, Thần tích Thần sắc, sách Hán Nôm viết tay,...

**5. Số hóa tài liệu là nghiệp vụ thư viện ứng dụng chủ yếu những thành tựu của công nghệ thông tin.** Sự lỗi thời nhanh chóng của phần mềm, thiết bị công nghệ thông tin sẽ tác động trực tiếp đến hoạt động số hóa. Do đó, người làm công tác số hóa phải tích cực nghiên cứu và nắm bắt kịp thời những

chuyển giao kỹ thuật và thay thế công nghệ để không bị lạc hậu so với sự phát triển chung của thế giới.

Công tác đào tạo cán bộ số hóa và tạo những cơ hội học hỏi kinh nghiệm, kỹ thuật tiên tiến về số hóa tài liệu được xem là yêu cầu quan trọng trong phát triển số hóa của tất cả các thư viện hiện đại.

Tài liệu số là một trong bốn thành tố quan trọng của thư viện số - thư viện thông minh. Hoàn thiện và tiếp tục phát triển số hóa tài liệu là vấn đề cần quan tâm của mỗi thư viện. Với điều kiện của Thư viện KHXH, số hóa

tài liệu cần được hoàn thiện thêm ở các vấn đề như:

- Bổ sung license phần mềm xử lý ảnh Scangate theo hệ thống thiết bị số hóa tự động, đảm bảo chất lượng ảnh số đạt tiêu chuẩn cao nhất;

- Đầu tư hệ thống lưu trữ dành riêng cho bộ phận bảo quản tài liệu số, hệ thống có khả năng sao lưu và xuất dữ liệu khi cần.

- Đầu tư phần mềm quản trị tài nguyên số, phần mềm có khả năng quản trị, cập nhật, trao đổi, đáp ứng nhu cầu tài liệu online của độc giả trên toàn thế giới □