

“SỐ HÓA TÀI LIỆU” - TỪ NHẬN THỨC ĐẾN TRIỂN KHAI ĐÀO TẠO TẠI KHOA THÔNG TIN-THƯ VIỆN, TRƯỜNG ĐẠI HỌC KHOA HỌC XÃ HỘI & NHÂN

*PGS.TS. Trần Thị Quý - Chủ nhiệm Khoa
Khoa Thông tin – Thư viện - Trường ĐH KHXH & Nhân văn, ĐHQG HN*

1. Nhận thức về “số hóa tài liệu”

+ Nội hàm khái niệm “số hóa tài liệu”

Ngày nay, sự phát triển mạnh mẽ của công nghiệp công nghệ thông tin đã tạo ra hàng loạt các sản phẩm công nghệ phần cứng, công nghệ phần mềm và công nghệ nội dung. Các loại hình sản phẩm này đã tác động và làm biến đổi về chất các loại hình thư viện truyền thống. Dần hình thành loại hình “Thư viện số”/“Thư viện điện tử (electronic library)/ “Thư viện ảo” (virtual library)/ “Thư viện không tường” (library without walls),... Một trong các yếu tố để xây dựng, duy trì và phát triển loại hình thư viện này là nguồn tài liệu số/tài nguyên tri thức dạng số. Có thể nói nguồn tài nguyên thông tin số là huyết mạch/linh hồn của thư viện số. Mục tiêu của bất kỳ Thư viện số nào là tạo ra một cổng thông tin truy cập trực tuyến đến tài nguyên số không chỉ của thư viện đó mà đến bất kỳ Thư viện số nào khác ở bất cứ đâu. Do đó, việc triển khai xây dựng Bộ sưu tập tài nguyên thông tin số là bước đi đầu tiên, quan trọng nhất để phát triển thư viện số. Khi các thư viện đã có Bộ sưu tập số sẽ tiến hành liên thông chia sẻ phục vụ cộng đồng người dùng tin trực tuyến.

Khi nói về thư viện số và nguồn tài nguyên tri thức số, ngay từ năm 1998, Liên đoàn thư viện số (Digital Libraries Federation - DLF) đã khẳng định: *“Các thư viện số là các tổ chức cung cấp nguồn lực, trong đó bao gồm các chuyên gia (những người có kiến thức và kỹ năng làm việc trong môi trường số), để lựa chọn thông tin, cấu trúc hóa, đưa ra các phương thức truy cập và phân phối thông tin hiệu quả, cũng như đảm bảo sự toàn vẹn và của bộ sưu tập số, sao cho chúng luôn sẵn sàng và kinh tế để phục vụ một cộng đồng cụ thể hoặc một nhóm cộng đồng.*

Để có Bộ sưu tập số hay nói cách khác muốn tạo lập, phát triển kho tài nguyên tri thức dạng số có 03 cách:

- Cách thứ nhất: Mua tài nguyên thông tin điện tử từ các nhà cung cấp/xuất bản/cá nhân (trước khi in ra giấy) hoặc trao đổi với các đối tác
- Cách thứ hai: Truy cập, khai thác từ việc liên kết đến các nguồn tài liệu số có cùng nội dung thông tin/thông tin chuyên biệt (thông tin Y học, thông tin về Luật học) trên Internet
- Cách thứ ba: Tổ chức số hóa nguồn tài nguyên thông tin truyền thống bằng phương pháp quét hay nhập lại thông tin từ bản phím của máy tính điện tử.

Trong ba cách trên, cách thứ ba là quan trọng nhất bởi lẽ: Mỗi một cơ quan thông tin, thư viện đều có thể mạnh riêng/bản sắc riêng nhờ vào đặc điểm vốn tài liệu mình đang lưu giữ. Đặc biệt là đặc điểm về nội dung của các loại hình tài liệu (tài liệu quý,

hiếm, có giá trị chất xám cao). Ắt ẽ số hóa được nguồn tài liệu này, xây dựng được bộ sưu tập số là vốn quý để duy trì và phát triển thư viện của mình. Từ đây, nhiều người của cộng đồng truy cập sẽ biết đến để chia sẻ khai thác nhiều lần, sử dụng nhiều lần.

Vậy, số hóa tài liệu là gì? đây là khái niệm vẫn còn mới đối với cộng đồng những người làm công tác thông tin, thư viện. Có nhiều định nghĩa khác nhau nhưng có nội dung chung đều cho rằng: *số hóa tài liệu là quá trình chuyển các dạng dữ liệu/tài liệu truyền thống như các bản viết tay, bản in trên giấy, hình ảnh, âm thanh, phim, dữ liệu toàn văn với nhiều định dạng khác nhau sang chuẩn dữ liệu trên máy tính và được máy tính nhận biết được như tài liệu ban đầu gọi là số hoá dữ liệu. Hay nói cách khác số hoá tài liệu là hình thức chuyển đổi các tài liệu truyền thống bên ngoài thành dạng tài liệu số mà máy tính có thể hiểu được. Sản phẩm sau khi số hóa tài liệu chính là nguồn Tài nguyên số/dữ liệu số - các dữ liệu dạng chữ, hình ảnh, âm thanh... được máy tính nhận biết đúng định dạng và được sử dụng trên máy tính.*

+ Vai trò của công tác số hóa tài liệu

Một khi nguồn tài nguyên thông tin số là mạch huyết, là linh hồn của thư viện số thì công tác số hóa tài liệu có vai trò vô cùng quan trọng. Trước hết việc số hóa tài liệu sẽ giúp giảm thiểu đáng kể diện tích, không gian lưu giữ; Thứ hai là giúp việc bảo quản, duy trì tuổi thọ của tài liệu truyền thống được lâu hơn; Thứ ba là dễ dàng mở rộng phạm vi cộng đồng người sử dụng nguồn tài nguyên thông tin của cơ quan thông tin, thư viện; Thứ tư là tiện ích trong việc truy xuất tìm kiếm thông tin ở bất kỳ đâu vào bất cứ thời điểm nào một cách nhanh chóng, dễ dàng; Thứ năm là thuận lợi trong việc chia sẻ nguồn tài nguyên thông tin của thư viện với các thư viện khác. Thứ sáu là giảm thiểu tối đa sức người, sức của cho việc quản lý nguồn tài nguyên thông tin truyền thống; Thứ bảy là góp phần nhanh chóng tái tạo thông tin mới có giá trị gia tăng cao hơn.

+ Các yếu tố liên quan đến hoạt động Số hóa tài liệu

Để tiến hành số hóa tài liệu xây dựng kho tài nguyên thông tin số cần phải chú trọng đến các yếu tố như: *Xác định mục tiêu số hóa tài liệu; Vấn đề công nghệ/lựa chọn phần mềm quản lý; Lựa chọn tài liệu để số hóa; Thực hiện Quy trình số hoá và xử lý tài liệu sau số hóa chặt chẽ; Nguồn nhân lực phục vụ số hóa tài liệu; Kinh phí số hóa tài liệu; Vấn đề đảm bảo bản quyền...*

- *Xác định mục tiêu số hóa tài liệu:* là công việc đầu tiên khi triển khai số hóa tài liệu. Số hóa tài liệu để phục vụ nghiên cứu, đào tạo hay kinh doanh; Phục vụ đối tượng cộng đồng người dùng tin nào...

- *Chú trọng lựa chọn các trang thiết bị chuyên dụng và Phần mềm quản lý tài liệu số với các tiêu chuẩn công nghệ đảm bảo chất lượng tài liệu số và quản lý tài liệu số bền vững.* Phần mềm này cho phép lưu trữ và khai thác các loại tài liệu đã được số hoá âm thanh, hình ảnh, phim, dữ liệu toàn văn với nhiều định dạng khác nhau. Hỗ trợ xây dựng các giáo trình, bài giảng, sách điện tử, biến các loại tài liệu này trở thành các dữ liệu của thư viện. Cho phép bạn đọc truy cập và khai thác thông tin trực tuyến. Các thành phần chính của phần mềm quản lý tài liệu số gồm có: **OPAC:** Cung cấp giao

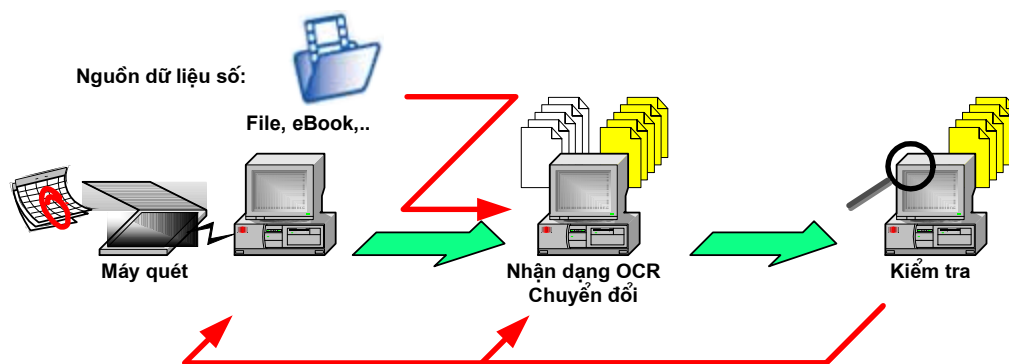
diện cho người dùng, qua đó truy cập đến các chức năng của hệ thống thư viện số như tra cứu, xem tài liệu...; **Library Server:** Tích hợp với hệ thống thư viện điện tử tích hợp, module này cung cấp các giao diện để truy cập thông tin bạn đọc, bản ghi biên mục...; **Object Server:** là nơi lưu trữ và cung cấp nội dung tư liệu; **Authority Control:** Có chức năng xác thực, kiểm soát và ghi nhận các truy cập hệ thống. Từ đó có thể đưa ra các báo cáo thống kê, thu phí,..

Phần mềm có các tính năng cơ bản sau: **Thu thập và bổ sung các tư liệu:** cung cấp quy trình số hóa và xử lý các dạng tài liệu khác nhau như văn bản, hình ảnh, âm thanh, phim và các định dạng khác; **Lưu trữ:** Hệ thống Thư viện Số cho phép lưu trữ các tư liệu thuộc nhiều dạng khác nhau: văn bản, hình ảnh, âm thanh; **Biên mục theo chuẩn:** ả nguyên tắc chung của việc biên mục dữ liệu số là xây dựng các bộ nhãn trường cá biệt cho mỗi loại tài liệu cần số hoá. Các nhãn trường này tuân theo các thành phần do tổ chức Sáng kiến siêu dữ liệu Dublin Core (DCMI) qui định. Tuân theo chuẩn RDF (Resource Description Framework) của W3C. Các bản ghi thư mục mô tả các nguồn tư liệu số hoá có thể được thể hiện dưới nhiều khuôn dạng khác nhau: MARC21, CDS/ISIS. ả ngoài ra, các bản ghi này có thể thể hiện dưới dạng các tệp XML, tuân theo Resource Description Framework (RDF) do tổ chức W3C khuyến cáo; **Tự xác định các thuộc tính của các tệp dữ liệu số hoá** được đưa vào để cho phép tra cứu theo các thuộc tính đó, ví dụ kích cỡ tệp, loại nén (với hình ảnh, âm thanh và video), cỡ và độ sâu (đối với hình ảnh), độ dài (đối với âm thanh và video)...; **Tra cứu Tìm kiếm:** Hệ thống cho phép tìm kiếm theo tổ hợp các thông tin mô tả đồng thời với các thuộc tính số. Ví dụ: Với các text, tệp, hệ thống có khả năng đánh chỉ mục tất cả các dạng tệp thông dụng (text, html, xml, word, excel, pdf,...) và cung cấp khả năng tìm kiếm toàn văn. Với hình ảnh, hệ thống Thư viện Số cho phép người dùng tìm kiếm các thuộc tính số của ảnh: kích thước, định dạng, độ lớn. Hơn nữa, Thư viện Số còn cung cấp khả năng tìm kiếm theo nội dung (content search) dựa trên các thông tin về phân bố màu sắc, bố cục, nền ảnh. ả người dùng có thể tìm kiếm các tư liệu âm thanh theo độ dài, mô tả, định dạng, số kênh; **Quản lý quyền truy cập, kiểm soát truy cập, nhật ký truy cập dữ liệu:** thư viện Số sử dụng chung hệ thống người dùng của phân hệ quản trị thư viện truyền thống, cho phép quản lý tập trung, mỗi người dùng chỉ có một tài khoản. Quyền truy cập được xác định trên nhóm tài liệu hoặc từng tài liệu. Quyền có thể gán cho nhóm người dùng hoặc từng người dùng. Các truy cập đều được ghi lại trong nhật ký hệ thống. Từ đó dễ dàng cung cấp các chức năng báo cáo thống kê hoặc tính phí.

- Về *lựa chọn tài liệu*: Là việc hết sức quan trọng, bởi không có thư viện nào có khả năng số hoá toàn bộ kho tài liệu. Khi lựa chọn tài liệu trước hết cần chú trọng đến nhu cầu thông tin (nội dung, ngôn ngữ, dạng tài liệu...) của người dùng tin mà thư viện đang phục vụ. Đặc biệt là nhu cầu về nội dung tài liệu/giá trị tri thức của tài liệu gốc; Thứ hai là cần tuyển lựa các tài liệu tiềm năng/đặc thù mà thư viện mình đang lưu giữ, các thư viện khác không có; Thứ ba là tính tới các tài liệu chỉ có một bản, tài liệu sắp

hư hỏng khó hồi phục; Thứ tư là tài liệu quý hiếm; Thứ năm là tài liệu chưa có nơi nào số hóa để tránh trùng lặp...

- *Thực hiện Quy trình Số hoá và xử lý các tài liệu sau số hóa chặt chẽ:* Số hoá và xử lý tài liệu sau số hóa là công đoạn quan trọng nhất để tạo nên một thư viện số nhờ quá trình tạo ra thông tin cho thư viện. Có rất nhiều các dạng tài liệu gốc (ví dụ như fulltext, video, audio, ảnh,...). Tài liệu có thể bao gồm cả chữ và hình ảnh, video bao gồm cả audio và hình ảnh,... Với mỗi dạng tài liệu đều có các cách xử lý khác nhau. Ắt hẳn nói chung đều phải qua các công đoạn số hoá (tạo ra các hình ảnh số) và sau đó là xử lý để tạo ra các thông tin số. Các thông tin số này mới thực sự là đối tượng của thư viện điện tử, cho phép thực hiện thao tác sửa đổi. Công ty máy tính truyền thông CMC mô hình hóa Quy trình số hoá dữ liệu toàn văn như sau:



Để thấy rõ sản phẩm của Quá trình số hóa ta có thể so sánh sự khác nhau giữa tài liệu in/tài liệu gốc và tài nguyên thông tin số:

	Bản gốc	Hình ảnh số	Thông tin số
Dạng vật chất	Vật thể vật chất (sách, video,...)	Tệp máy tính	Tệp máy tính
Định dạng (format)	Đa dạng (văn bản tiếng Anh, VHS, ...)	Tệp đồ họa (.BMP, .MPG,...)	Tệp có cấu trúc (.DOC, .MPG) biểu ghi chỉ số và cơ sở dữ liệu
Khả năng đọc	ả gười hoặc là các thiết bị chuyên dụng	Các chương trình đồ họa máy tính	Các chương trình máy tính như văn bản, video hay cơ sở dữ liệu
Khả năng nhân bản	ả hân bản vật chất (photocopy,...)	Copy tệp và in các bản sao y hệt theo số lượng mong muốn	Tạo các thông tin của bản gốc theo các định dạng khác nhau (in lại sách theo chữ

	Bản gốc	Hình ảnh số	Thông tin số
			in nghiêng, chơi video với các track âm thanh khác nhau
Xử lý	Thay đổi bằng tay (viết vào lề sách, cắt và nối băng)	Đánh dấu bằng máy tính và xử lý đồ hoạ (thêm vào các ghi chú của người dùng, phóng to/thu nhỏ, thay đổi màu sắc, thay bằng các hình ảnh khác.	Sửa đổi lại các thông tin gốc, tạo ra các tài liệu mới từ tài liệu gốc, sao chép và phân phối không giới hạn.

- *Về nhân lực phục vụ số hóa tài liệu:* Trước khi tiến hành số hóa tài liệu cần chú trọng tuyển lựa những người có khả năng tiếp nhận và làm chủ công nghệ số hóa tài liệu, đảm bảo tài liệu được số hóa có chất lượng và quản trị tài nguyên thông tin số từ nhà cung cấp thiết bị và công nghệ.

- *Về Kinh phí số hóa tài liệu:* Số hóa tài liệu là hoạt động cần một khoản kinh phí khá lớn, đòi hỏi cần cân đối chi phí giữa việc mua các trang thiết bị scan chuyên dụng với các phần mềm nhận dạng đảm bảo tính bền vững từ khâu phát triển, lưu giữ, bảo quản đến khai thác và kinh phí để triển khai đào tạo nguồn nhân lực số hóa tài liệu, cũng như người dùng tin để họ biết cách sử dụng hệ thống thông tin...

- *Về vấn đề đảm bảo bản quyền:* khi số hóa tài liệu cần chú trọng đến các văn bản pháp quy liên quan đến quyền tác giả như Công ước BERNE (những tài liệu đã xuất bản trên 50 năm thì được quyền số hóa). Ở Việt Nam, vấn đề bản quyền đã được đề cập đến trong Bộ Luật Dân sự năm 2005, Luật Sở hữu trí tuệ và Nghị Định 100/NĐ-CP/2006 của Chính phủ quy định chi tiết và hướng dẫn thi hành một số điều của Bộ luật Dân sự, Luật Sở hữu trí tuệ về quyền tác giả và quyền liên quan.

2. Triển khai đào tạo tại Khoa Thông tin-Thư viện

Ở Việt Nam, việc tiếp cận thư viện số và xây dựng bộ sưu tập số cho thư viện mới chỉ ở bước đầu. Trong chương trình đào tạo tại Khoa TT-TV, trường ĐHKHXH&V, “Số hóa tài liệu” chưa có tên gọi trực tiếp và thiết kế thành một chuyên đề riêng biệt. Mặc dù vậy, cách đây 6 năm, khi xây dựng chương trình đào tạo theo phương thức tín chỉ, chúng tôi đã thiết kế môn học “Công nghệ nội dung” và đưa vào chương trình đào tạo. Đây là môn học bắt buộc và có thời lượng 2 tín chỉ, tương đương với 3 đơn vị học trình của phương thức đào tạo theo niên chế. Mục tiêu của môn học “Công nghệ nội dung” chúng tôi đã xác định là sẽ trang bị cho sinh viên:

Về kiến thức:

- ă ăm được nội hàm các khái niệm liên quan tới “công nghệ nội dung” và “công nghiệp nội dung”.
- ă ăm được đặc điểm của các dạng tài liệu điện tử/tài liệu số.
- Hiểu được vai trò và lợi ích của tài liệu số trong thời đại ngày nay.
- ă ăm được các phương pháp tạo lập, khai thác, lưu trữ và bảo quản tài liệu điện tử/tài liệu số.
- ă ăm được các vấn đề liên quan tới an toàn và bảo mật các tài liệu điện tử/tài liệu số;
- Hiểu được các vấn đề liên quan tới bản quyền của tài liệu điện tử.
- Biết và hiểu được chức năng và công dụng của một số công cụ và phần mềm xử lý và số hóa các tài liệu ở các kiểu định dạng file khác nhau.
- ă ăm được một số ngôn ngữ siêu văn bản.
- Hiểu được một số công cụ và phần mềm liên quan tới Web và Internet.
- ă ăm được tính chất và tầm quan trọng của việc ứng dụng công nghệ nội dung trong các lĩnh vực chính trị, kinh tế, văn hóa, giáo dục, giải trí, đặc biệt là lĩnh vực thông tin – thư viện.

Về kỹ năng:

- Sử dụng thành thạo một số công cụ và phần mềm xử lý các kiểu định dạng của tài liệu số
- Biết số hóa các dạng tài liệu.
- Biết tạo lập, khai thác, lưu trữ và bảo quản tài liệu số.
- Biết sử dụng ngôn ngữ siêu văn bản để xây dựng một trang Web cá nhân trong đó đưa vào các loại định dạng cơ bản của tài liệu số như: âm thanh, hình ảnh, phim (tài liệu đa phương tiện), toàn văn.

Về thái độ:

- Có ý thức quan tâm tới xu hướng phát triển của công nghệ nội dung và công nghiệp nội dung trên thế giới và ở Việt ă ăm.
- Có ý thức trong việc nâng cao trình độ tin học ứng dụng nói chung.
- Thích môn học, muốn nghiên cứu tìm tòi, cập nhật và rèn luyện kỹ năng sử dụng các công nghệ số hóa tài liệu.

Bên cạnh chuyên đề “Công nghệ nội dung” đó, trong chương trình đào tạo cử nhân vấn đề Số hóa tài liệu còn được đề cập đến ở một số các chuyên đề/môn học khác như: ă ăm hàm khái niệm Tài liệu số, Bộ sưu tập số trong môn “Thông tin học đại cương”; Vấn đề Bản quyền trong môn học “Phát triển vốn tài liệu”; Phần mềm quản trị tài liệu, số hóa và xử lý tài liệu trong môn “Tự động hóa công tác TTTV; Tài liệu số trong môn học “Thư viện điện tử”; Các chuẩn liên quan đến phát triển sưu tập số trong môn học “Biên mục tự động” và môn học “Các phần mềm quản trị hoạt động TTTV”....

Tuy nhiên, khi triển khai đào tạo chúng tôi gặp không ít khó khăn: môn học “Công nghệ nội dung” là môn học mới, giáo trình chưa có, tài liệu tham khảo chưa nhiều; Các trang thiết bị và phần mềm chưa có vì vậy, việc thực hành của sinh viên để củng cố kiến thức, rèn luyện kỹ năng bị hạn chế.

Hiện nay, chúng tôi bước vào giai đoạn 2 của việc chuyển đổi phương thức đào tạo theo tín chỉ theo chiều sâu, cấu trúc lại nội dung, chương trình đào tạo cho cả ba cấp đào tạo: cử nhân, thạc sĩ và tiến sĩ vì vậy, đây là cơ hội rất thuận lợi để chúng tôi nghiên cứu một cách nghiêm túc, thiết kế đưa nội dung Số hóa tài liệu thành một chuyên đề độc lập với tên gọi phù hợp, có thể là “Phát triển nguồn tài nguyên thông tin số”./.

TÀI LIỆU THAM KHẢO

1. Cao Minh Kiểm. Về các chuẩn áp dụng trong số hóa tài liệu phục vụ xây dựng thư viện điện tử và trao đổi dữ liệu//Kỷ yếu Hội thảo khoa học chuyên đề: quản trị và chia sẻ nguồn tin số hóa.H, 2007.
2. Nguyễn Tiến Đức.Ứng dụng phần mềm nguồn mở Zope để xây dựng cổng thông tin về khoa học và công nghệ ở địa phương//Kỷ yếu Hội thảo khoa học chuyên đề “Phát triển thư viện số ở Việt ấ am”.H, 2007.
3. Tìm hiểu Luật công nghệ thông tin.H.: Lao động, 2007.-131 tr.
4. Trần Thị Quý. Tự động hóa trong hoạt động thông tin-thư viện/Trần Thị Quý, Đỗ Văn Hùng.H.: ĐHQG Hà , 2007.- 270 tr.
5. Viện ấ ngôn ngữ học.Từ điển tiếng Việt.H.: Đà ấ ẵng, 2006.-1170 tr.
6. Grace, S. W. K. (2009). Digital libraries overview and globalization. In S. F. Yin-Lang Theng, Dion Goh, Jin-Cheon ấ a (Ed.), *Handbook of research on digital libraries: design, development and impact* (pp. 562-573). ấ Y.: Information science reference.
7. Lynch, C. (2002). Digital collections, digital libraries and digitization of cultural heritage information. *First Monday*, 7(5-6).