

## TÀI LIỆU NỘI SINH – NGUỒN TIN VÔ GIÁ TRONG CÁC TRƯỜNG ĐẠI HỌC

**TS. HUỖNH MÃN ĐẠT**

Trường ĐH Văn hóa Tp. Hồ Chí Minh

Tel: 0888 969 678, mandat77@yahoo.com

**Tóm tắt:** Bài viết trình bày có hệ thống các khái niệm về tài liệu nội sinh, đưa ra quy trình số hóa nguồn tài liệu nội sinh, lựa chọn công nghệ và đưa ra các giải pháp nhằm giúp các trường đại học làm tốt công tác số hóa nguồn tài liệu nội sinh.

Tài liệu nội sinh là những tài liệu được hình thành trong quá trình hoạt động khoa học kỹ thuật, sản xuất, quản lý, nghiên cứu, học tập của các cơ quan, tổ chức, xí nghiệp, viện nghiên cứu, trường đại học...

Tài liệu nội sinh phản ánh đầy đủ, có hệ thống về các thành tựu, tiềm lực cũng như hướng phát triển của những đơn vị này và thường được lưu giữ ở các thư viện và trung tâm thông tin của đơn vị đó.

Đặc điểm của nguồn tài liệu nội sinh:

Theo tính chất của quá trình tạo ra nguồn tài liệu nội sinh có thể chia thành 3 nhóm:

- Nguồn tin phản ánh các kết quả hoạt động đào tạo: Luận án, luận văn, các kết luận khoa học, các tư liệu điền dã, các tư liệu điều tra, các hồ sơ thí nghiệm, các chương trình đào tạo, giáo trình, đề cương bài giảng...

- Nguồn tin phản ánh kết quả NCKH: Các báo cáo kết quả nghiên cứu khoa học, tư liệu trung gian được tạo nên từ việc triển khai các chương trình, đề tài NCKH, đề án, dự án sản xuất thử, báo cáo khoa học, kỷ yếu hội nghị, hội thảo...

- Nguồn tin phản ánh tiềm lực đào tạo và NCKH: Bao gồm các tài liệu về cơ cấu, quy mô, trình độ đào tạo, nguồn nhân lực khoa học, cơ sở vật chất kỹ thuật để triển khai các hoạt động đào tạo, NCKH, các thông tin phản ánh định hướng phát triển của nhà trường.

• Lợi ích đối với cơ quan: Đưa ra một chỉ số tốt về chất lượng và hiệu suất của cơ quan, tăng cường hình ảnh và uy tín, tăng cường khả năng tiếp cận đối với kết quả nghiên cứu, nội dung với chất lượng cao có thể sẽ là công cụ quảng bá để thu hút nhân sự, sinh viên và các nguồn đầu tư.

• Lợi ích đối với nhà nghiên cứu: Mở rộng việc phát tán và cung cấp các công bố của mình, gia tăng tác động của các công bố (các nghiên cứu được tự do tiếp cận sẽ dễ dàng được trích dẫn), đưa ra cách đo lường để các nhà nghiên cứu có thể xác định tỉ lệ truy cập đối với từng nghiên cứu cụ thể, giúp quản lý và lưu giữ các nội dung liên quan đến nghiên cứu của từng cá nhân, cho phép tạo ra các danh mục xuất bản phẩm theo yêu cầu cá nhân.

• Lợi ích đối với cộng đồng quốc tế: Hỗ trợ hợp tác nghiên cứu thông qua việc tạo điều kiện trao đổi tự do cho các nguồn thông tin học thuật, giúp cộng đồng hiểu về

các nỗ lực và các hoạt động nghiên cứu, lợi ích đối với cán bộ thư viện, luôn phù hợp trong thời đại số với nhiều thay đổi và tiến triển, là cơ hội để thể hiện vai trò quan trọng hơn trong bối cảnh thay đổi của việc truyền tải thông tin học thuật.

Tài liệu nội sinh bao gồm cả tài liệu đã xuất bản lẫn tài liệu chưa xuất bản. Cụ thể như sau:

Nhóm tài liệu đã xuất bản: bài báo được đăng trên các báo, tạp chí; sách; tài liệu hội nghị hội thảo.

Nhóm tài liệu chưa xuất bản: bản tài liệu trước khi in; các công trình chưa công bố hoặc phần nội dung được công bố của các công trình chưa hoàn tất, luận văn, luận án, báo cáo khoa học.

Nhóm tài liệu hỗ trợ học tập và giảng dạy: đề cương, giáo án, bài giảng, ngân hàng đề thi, băng hình phục vụ các khóa học.

Như vậy, nguồn nội sinh tốt sẽ mang đến nhiều lợi ích cho các thành viên liên quan; do đó các thư viện cần tuyên truyền những lợi ích này để mọi người sử dụng ủng hộ để phát triển nguồn thông tin nội sinh.

Quy trình số hóa nguồn tài liệu nội sinh có thể chia thành 5 bước cụ thể sau:

#### **Lựa chọn tài liệu đầu vào:**

Đây là công đoạn đầu tiên trong quy trình số hóa tài liệu, bao gồm việc cân nhắc, lựa chọn và xác định những đối tượng tài liệu nào được đưa vào số hóa. Các thư viện cần xây dựng chính sách thu thập đối với tài liệu nội sinh ngay từ ban đầu là rất quan trọng nhằm đảm bảo tổ chức và sử dụng được một nguồn tài liệu học thuật một cách đúng đắn, cũng như đảm bảo được sự phát triển ổn định nguồn tài liệu này. Chính sách phải được nêu lên những nội dung cơ bản sau: xác định các loại tài liệu cần thu thập; đối tượng nộp và đối tượng sử dụng nguồn nội sinh; mức độ phổ biến và mức độ cho phép sử dụng nguồn nội sinh, chính sách bảo quản nguồn nội sinh, chính sách cập nhật nguồn nội sinh.

Về bản quyền của tài liệu: trong trường hợp thư viện cung cấp truy cập mở cho nguồn nội sinh thì các tác giả và bộ phận tiếp nhận tài liệu cần làm rõ vấn đề pháp lý liên quan đến bản quyền để bảo đảm thư viện không vi phạm luật bản quyền.

Nội dung tài liệu: trên cơ sở xác định nhóm người dùng tin (Cán bộ lãnh đạo, giảng viên, cán bộ nghiên cứu, sinh viên, các đối tượng khác,...), mục tiêu, chức năng nhiệm vụ của thư viện mà thư viện lựa chọn các tài liệu nội sinh có nội dung phù hợp, tài liệu có tần suất sử dụng cao.

Điều kiện bảo quản hiện tại: tùy tình hình cụ thể của từng thư viện trong tình trạng điều kiện bảo quản kết hợp với nội dung tài liệu nội sinh mà quyết định lựa chọn tài liệu để tiến hành số hóa. Ưu tiên số hóa các tài liệu nội sinh mà nhu cầu sử dụng của người dùng cao.

#### **Lựa chọn công nghệ:**

Lựa chọn công nghệ để tiến hành số hóa tài liệu nội sinh đóng vai trò rất quan trọng bởi đây là công cụ đắc lực giúp các trường đại học thực hiện các công việc trong quy trình tạo lập và vận hành bộ sưu tập số, công nghệ để tiến hành số hóa cần đáp ứng các yêu cầu sau:

- Là công cụ, môi trường để đảm bảo các tài liệu số hóa sau khi được tạo lập sẽ dễ dàng, thuận tiện cho người dùng tiếp cận;

- Có đủ độ tin cậy cho người quản trị và kỹ thuật viên trong quá trình tạo lập, bảo quản và cung cấp dữ liệu trong quá trình hoạt động của bộ sưu tập;

- Đáp ứng đầy đủ các yêu cầu về chuẩn nghiệp vụ thông tin – thư viện;

- Dễ dàng trao đổi dữ liệu với các chuẩn khác, có công cụ sao lưu an toàn dữ liệu với các chuẩn khác, có công cụ lưu sao an toàn dữ liệu.

Để bộ sưu tập số phát huy được hết tác dụng, thư viện khi thực hiện tạo lập bộ sưu tập số cần phải có cơ sở hạ tầng sau:

- Hệ thống mạng intranet được kết nối internet với đường truyền đủ đáp ứng cho số người dùng tối thiểu của thư viện;

- Hệ thống máy chủ đủ mạnh để đáp ứng việc lưu trữ, bảo quản, cung cấp dữ liệu và quản lý người dùng và các phần mềm hệ thống có bản quyền;

- Trang web đăng tải và là cổng truy cập của người dùng vào bộ sưu tập.

### **Số hoá nguồn tài liệu:**

Đây là công đoạn đòi hỏi đầu tư nhiều công sức, kinh phí nhưng lại là khâu dễ dàng thực hiện nhất.

Việc nộp tài liệu vào nguồn nội sinh có thể thực hiện theo 2 cách: trực tuyến hoặc thông qua cán bộ thư viện (nộp bản giấy). Trong trường hợp thư viện đã áp dụng công nghệ chuẩn bị sẵn 1 giao diện nộp tài liệu nội sinh trên website, tác giả có thể tự nộp trực tuyến. Trường hợp chúng ta thu thập được hoặc tác giả cung cấp tài liệu giấy, 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.

**Biên mục tài liệu số hóa (Tạo siêu dữ liệu liên kết):** Mô tả dữ liệu (theo một trong các chuẩn siêu dữ liệu MARC, Dublin Core, MODS, METS, ISO 2709 trong đó chuẩn Dublin Core tương đối phổ biến vì có khả năng tùy biến cho các tiêu chuẩn khác với 15 trường biên mục);

Có nhiều chuẩn biên mục mang tính chất siêu dữ liệu khá thông dụng như: MARC 21/ UNIMARC, Dublin Core Metadata, XML... Các dữ liệu này thường được gắn vào phần đầu cho mỗi tài liệu điện tử đặt trên website và rất thích hợp cho các máy tìm kiếm, lọc ra thông tin để tổ chức thành kho dữ liệu mà không cần dùng đến hệ quản trị cơ sở dữ liệu truyền thống.

Dublin Core Metadata là chuẩn dùng mô tả nội dung của biểu ghi và dữ liệu. Nó đơn giản hơn MARC Format vì chỉ có 15 trường: nhan đề, tác giả, chủ đề, mô tả, nhà xuất bản, tác giả phụ, ngày tháng, loại tài liệu, mô tả vật lý, định danh, nguồn gốc, ngôn ngữ, liên kết, bao quát, bản quyền (trong khi MARC có đến hơn 200 trường, khá phức tạp).

Siêu dữ liệu (metadata) dùng để mô tả một tài nguyên thông tin được chia sẻ trên internet. Một bản ghi siêu dữ liệu bao gồm một tập hợp các thuộc tính hoặc tập các phần tử cần thiết để mô tả các tài nguyên theo yêu cầu. Tạo siêu dữ liệu theo 3 dạng (siêu dữ liệu mô tả: mô tả các thông tin về tài liệu, siêu dữ liệu cấu trúc: mô tả các liên kết giữa các đối tượng thông tin liên quan của tài liệu như mục lục, chương, phần, trang sách, hình ảnh minh họa, phụ lục... giúp người dùng dễ dàng di chuyển đến các thành phần của tài liệu, siêu dữ liệu quản trị: gồm tạo kích cỡ tập tin, định dạng tài liệu (PDF), đặc tính sử dụng và tình trạng của tài liệu).

- Siêu dữ liệu kỹ thuật: Thông tin về máy và sự vận hành trong quá trình chụp hình ảnh và thông tin này được tạo ra tự động bởi hệ thống của thư viện.

- Siêu dữ liệu cấu trúc: Thông tin về cấu trúc sách/trình tự sắp xếp đòi hỏi nhập liệu bằng tay.

- Siêu dữ liệu mô tả: Thông tin về cuốn sách là thông tin dưới biểu ghi MARC tương thích hoàn toàn tiêu chuẩn biên mục dữ liệu điện tử Dublin Core 2. Dữ liệu biểu ghi MARC được nhập với khả năng đọc số ISBN bằng mã số mã vạch (Barcode) hoặc một giao diện người dùng dành cho nhập liệu mô tả nội dung (Vd., tên nhan đề, tác giả, ngày bản quyền, bảng nội dung,...) trong phần mềm biên mục nhằm nhập liệu nhanh và dễ dàng sử dụng.

**Vận hành, bảo quản và cung cấp dữ liệu:** Trước khi vận hành thật các công đoạn: quét (scan) – đối với các tài liệu là sách, biên mục tài liệu, tải tài liệu lên mạng,... thư viện sẽ thực hiện giai đoạn thử nghiệm bằng cách cho scan khoảng 10 đơn vị tài liệu với đủ các loại hình: sách, tạp chí, bản thảo, tài liệu hành chính, tranh ảnh,... và cho lưu trữ cũng như vận hành thử trên website để kiểm tra về chất lượng, bao gồm kích cỡ của hình ảnh, vấn đề xử lý chung, dạng tập tin, chiều sâu của bit, vùng sáng, vùng tối, giá trị âm thanh, độ sáng, độ tương phản, độ phân giải, sự nhiễu, sự định hướng, tiếng động, sự điều chỉnh kênh màu, sự mất văn bản, sự điều chỉnh hình ảnh, sự mất đường truyền hay mất hình ảnh, sự sống động, chất lượng truy cập, hình thức ngắn gọn, rõ ràng của văn bản...

Trong quá trình quét tài liệu, tạo sản phẩm số cho đến biên mục tài liệu số nên được sao lưu, cất giữ bảo quản ở các dạng: bộ nhớ lớn của máy chủ, trên CD-ROM, trên ổ cứng di động...

Cung cấp, tải dữ liệu lên mạng là khâu cuối cùng của tiến trình số hóa, bao gồm việc đưa bộ sưu tập lên mạng của thư viện để phục vụ trực tuyến và thiết kế giao diện với người dùng: tạo ra các công cụ sử dụng, chính sách khai thác đối với người dùng, ý kiến đóng góp, đánh giá của người sử dụng, xây dựng các ứng dụng tùy biến, chính sách phát triển nguồn tài liệu... Tất cả các kết quả này cần được thông qua trước hội đồng số hóa để hoàn chỉnh lần cuối trước khi công bố kết quả bộ sưu tập đối với người dùng tin.

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.

Để công tác bảo quản, khai thác sử dụng tài liệu 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; Như vậy chúng ta sẽ giảm đi công đoạn tốn kém số hóa từ tài liệu giấy sang tài liệu điện tử.

Để công việc số hóa tài liệu nội sinh thành công một nội dung quan trọng là nhiệm vụ của nhân viên thư viện- những người quản lý và phục vụ nguồn nội sinh trong trường đại học. Nhân viên thư viện là người soạn thảo chính sách quy định việc thu thập, quản lý và sử dụng cho nguồn nội sinh và thuyết phục các đối tượng liên quan chấp thuận và thực thi chính sách này. Nhân viên thư viện thiết lập mối quan hệ hợp tác với tác giả của nguồn nội sinh để khuyến khích và hỗ trợ họ trong việc nộp cũng như trong việc sử dụng. Nhân viên thư viện cần phải thành thạo các kỹ thuật, các thao tác sử dụng công nghệ, cũng như tự xử lý hoặc đưa ra hướng giải quyết cho các sự cố có thể xảy ra.

## TÀI LIỆU THAM KHẢO

1. Bùi Loan Thùy (2013). *Khả năng chia sẻ nguồn tài nguyên điện tử và vấn đề bản quyền trong thư viện trường đại học*, Kỷ yếu hội thảo “ chia sẻ nguồn lực thông tin điện tử trong hệ thống thư viện cao đẳng, đại học Việt Nam”.
2. Nguyễn Hữu Viêm (2004). *Sách điện tử: thách thức của phát triển*. Tạp chí Thông tin tư liệu, số 4/2004; tr.: 20-22.
3. Huỳnh Mẫn Đạt (2016). “OCLC- cầu nối thư viện Việt Nam ra thế giới”, Văn hóa và nguồn lực, số 5(2016):tr89-93
4. Phạm Thúc Trương Lương (2006). Vấn đề bản quyền tác giả trong kỷ nguyên số: góc nhìn từ thư viện. Kỷ yếu hội thảo “tăng cường công tác tiêu chuẩn hóa trong hoạt động thông tin tư liệu.
5. Sharples, M., Corlett, D., Westmancott, O. (2000). *The Design and Implementation of a Mobile Learning Resource*, UK: University of Birmingham, Edgbaston.
6. Хуинь Ман Дат (2010), *Корпоративная деятельность вузовских библиотек во Вьетнаме/ Хуинь Ман Дат*, Вестник МГУКИ, № 2, С. 143 – 146.

**Từ khóa:** *Thư viện trường đại học, số hóa tài liệu, tài liệu nội sinh.*

*The article presents the system of endogenous material conceptions, bringing out the digitization process of endogenous resources, selecting technology and offering solutions to help the universities fulfill their tasks of digitizing endogenous resources.*

**Keywords:** *University Library, digitized documents, internal documents.*