

# GIẢI PHÁP XÂY DỰNG BỘ SƯU TẬP SỐ HIỆU QUẢ

*ThS. Kiều Thúy Nga\**

*ThS. Lê Đức Thắng\*\**

## **Đặt vấn đề**

Số hóa tài liệu, xây dựng các bộ sưu tập số là xu hướng tất yếu của thư viện, nhưng nếu có dự án/chương trình số hóa, bạn sẽ thực hiện như thế nào? bắt đầu từ đâu? triển khai như thế nào?, làm thế nào để đảm bảo hiệu quả một chương trình số hóa thường rất tốn kém, nhiều rủi ro.... và còn rất nhiều câu hỏi khác nữa.

Bài viết đề xuất một số giải pháp nhằm xây dựng bộ sưu tập số hiệu quả trong hoạt động thư viện và lưu trữ tại Việt Nam.

## **1. Các nguyên tắc xây dựng bộ sưu tập số**

Tháng 12/2007, tổ chức Tổ chức Tiêu chuẩn thông tin quốc gia Hoa Kỳ (NISO) đã xuất bản tài liệu “Hướng dẫn xây dựng các bộ sưu tập số hiệu quả” (A Framework of Guidance for Building Good Digital Collections), tài liệu xuất bản lần thứ 3, một tài liệu hướng dẫn thực hành cho việc xây dựng các bộ sưu tập số.

Theo đó, quá trình xây dựng một bộ sưu tập số được chia thành bốn nhóm tác vụ chính:

- Các bộ sưu tập (tổ chức các nhóm đối tượng).
- Các đối tượng (các vật liệu số).
- Siêu dữ liệu (thông tin về các đối tượng và các bộ sưu tập).
- Các sáng kiến (các chương trình hoặc dự án để tạo lập và quản lý các bộ sưu tập).

Đồng thời NISO cũng đề xuất 9 nguyên tắc nhằm xây dựng các bộ sưu tập số hiệu quả, bao gồm:

*Một là:* Một bộ sưu tập số được tạo lập tuân thủ quy tắc phát triển bộ sưu tập rõ ràng đã được thống nhất trước khi việc tiến hành xây dựng bộ sưu tập.

*Hai là:* Bộ sưu tập cần mô tả sao cho người sử dụng có thể nắm bắt được những đặc trưng của bộ sưu tập đó, bao gồm phạm vi, dạng thức, mức độ được truy cập,

---

\* Thư viện Quốc gia Việt Nam

\*\* Thư viện Quốc gia Việt Nam

quyền sở hữu hoặc bất kỳ thông tin cần thiết nào để xác định nguồn gốc và sở hữu của bộ sưu tập đó.

*Ba là:* Một bộ sưu tập có chất lượng cao cần được quản lý tốt trong suốt vòng đời của nó. Quản lý số chú trọng đến việc quản lý vòng đời của tư liệu từ khi được tạo lập hoặc thu thập cho đến khi nó được loại bỏ. Việc quản lý bao gồm một loạt các tác vụ như quản trị dữ liệu tích cực, lưu giữ và bảo quản số.

*Bốn là:* Một bộ sưu tập tốt cần được phổ biến một cách rộng rãi và tránh những trở ngại cho người sử dụng tiếp cận. Quy tắc này bao gồm các thuộc tính: khả năng sẵn sàng, tính khả dụng, khả năng truy cập. Khả năng sẵn sàng nghĩa là bộ sưu tập có thể cho phép những người có quyền hạn truy cập và sử dụng. Điều này có nghĩa là những bộ sưu tập có thể truy cập thông qua nhiều hình thức, sử dụng những công nghệ phù hợp với những người sử dụng mục tiêu.

*Năm là:* Một bộ sưu tập tốt cần lưu ý đến những quyền sở hữu trí tuệ: phát triển bộ sưu tập tham khảo các quy định về bản quyền trong phạm vi Việt Nam (Luật Dân sự 2005; Luật Sở hữu trí tuệ 2009) cũng như quy định bản quyền trên phạm vi thế giới (Hiệp định Berne (Paris, 1971); Công ước bản quyền WIPO (Geneva, 1996); Công ước các bản ghi âm và các chương trình trình diễn WIPO (WPPT) (Geneva 1996).

*Sáu là:* Một bộ sưu tập tốt phải có những cơ chế thu thập dữ liệu để có thể đánh giá việc sử dụng và mức độ hữu dụng: Những các bộ sưu tập số cần được đánh giá định kỳ để theo dõi mức độ sử dụng, tính hiệu quả của dịch vụ. Việc quản lý bộ sưu tập hiệu quả cần đến nhiều phương pháp nghiên cứu để định lượng mức độ hữu dụng của bộ sưu tập, có thể thực hiện việc quan sát, lấy ý kiến, chú trọng đến những nhóm người sử dụng, phỏng vấn, thử nghiệm, nghiên cứu các tình huống và phân tích nhật ký phục vụ để đánh giá mức độ sử dụng và tính hữu dụng.

*Bảy là:* Một bộ sưu tập tốt phải có tính liên thông: Đảm bảo khả năng chia sẻ siêu dữ liệu (metadata) của chúng với những cơ chế tìm kiếm bên ngoài.

*Tám là:* Một bộ sưu tập tốt tích hợp vào chu trình công việc của người làm công tác thư viện và người sử dụng.

*Chín là:* Một bộ sưu tập tốt phải không bị lạc hậu: Những các bộ sưu tập số chứa những tư liệu có giá trị lâu dài nên được duy trì liên tục và lưu giữ để đảm bảo có thể truy xuất được. Tính liên tục cần được đặt ra từ các khía cạnh tổ chức, tài chính và kỹ thuật. Đặc biệt, những các bộ sưu tập số được xây dựng với ngân sách lớn cần có kế hoạch để tiếp tục việc duy trì, bảo dưỡng và hỗ trợ cả việc sau khi không còn ngân

sách hoạt động. Một cách tối ưu, bất kể nó được hình thành như thế nào, các bộ sưu tập số nên được tích hợp vào chu trình công việc quản lý những bộ sưu tập chung.

## **2. Các giải pháp cụ thể xây dựng bộ sưu tập số**

### **2.1. Hoàn thiện công tác lập kế hoạch**

Đối với công tác lập kế hoạch, cần thực hiện theo các bước:

Bước 1: Xác định mục tiêu và phạm vi của dự án, công việc

Bước 2: Xác định rõ đối tượng phục vụ theo từng nhóm cụ thể

Bước 3: Phân tích kho tài liệu để có thông tin cụ thể đối với từng dự án.

Bước 4: Lập tự toán chi tiết kinh phí, đồng thời phân sự tác động, ảnh hưởng của kinh phí khi thực hiện công việc này đến đơn vị.

Bước 5: Rà soát các tiêu chuẩn kỹ hoặc quy trình kỹ thuật đã có, nếu có cần áp dụng để đảm bảo tính thống nhất của dự án này và dự án khác.

### **2.2. Hoàn thiện công tác lựa chọn cho các bộ sưu tập số**

\* Xây dựng bộ tiêu chí lựa chọn tài liệu số hóa.

Câu hỏi là làm thế nào để có thể lựa chọn được tài liệu chính xác, cần thiết nhất để số hóa? Trong thực tế việc lựa chọn nội dung tài liệu để số hóa thường được thực hiện một cách chủ quan, trong nhiều trường hợp còn phụ thuộc vào ý chí chủ quan của lãnh đạo bộ phận hoặc trưởng nhóm triển khai, do đó nhiều khi nội dung tài liệu số hóa không phù hợp hoặc chưa thực sự phù hợp cho dự án đang/đã thực hiện.

Để khắc phục tình trạng này cần xác định và xây dựng bộ tiêu chí lựa chọn tài liệu số hóa, giúp công tác lựa chọn một cách khách quan hơn, đồng thời hỗ trợ lãnh đạo đơn vị, quản lý bộ phận/nhóm thực thi số hóa tài liệu ra quyết định một cách phù hợp. Cách đơn giản nhất là đặt nhiều câu hỏi trước khi thực hiện số hóa, xây dựng các bộ sưu tập số:

- Đối tượng phục vụ hoặc đối tượng người sử dụng sẽ hướng đến?

- Việc số hóa tài liệu có ảnh hưởng đến hoạt động của đơn vị hoặc các bộ phận khác không?

- Tình trạng bản quyền của tài liệu ra sao? Tài liệu còn thời hạn bảo hộ bản quyền không?, nếu còn thì tổ chức/cá nhân nào đang sở hữu để thực hiện các bước tiếp theo trong trường hợp tài liệu không thể thiếu đối với dự án như: giấy phép, tài chính chi trả, thỏa thuận sử dụng...

- Tình trạng vật lý của tài liệu như thế nào?, trong thực tế tài liệu thư viện rất đa dạng, phong phú với nhiều tình trạng tài liệu khác nhau, do đó đòi hỏi công tác lựa chọn cần lưu ý đến tình trạng vật lý của tài liệu để có thể xác định khối lượng công việc hoặc quy trình công việc (Ví dụ: phục chế trước khi số hóa), hoặc cả cân nhắc các giải pháp về kỹ thuật số hóa phù hợp (Đối với từng loại tài liệu khác nhau sẽ có phương án số hóa khác nhau với từng thiết bị phù hợp)

\* Ban hành quy định về tiêu chí lựa chọn tài liệu

Xây dựng quy định là một phần của công tác lựa chọn tài liệu để số hóa, xây dựng các bộ sưu tập, do đó cần thiết phải “*văn bản hóa*” các tiêu chí đang được sử dụng để lựa chọn tài liệu cho việc số hóa. Bởi việc này phục vụ nhiều mục đích. Đầu tiên, trong quá trình thực hiện, triển khai một dự án nào đó, nó cho phép bạn rà soát lại các tài liệu được lựa chọn để đảm bảo rằng việc lựa chọn tài liệu số hóa đã được áp dụng một cách nhất quán phù hợp với các tiêu chí đặt ra.

Mặt khác một văn bản quy định các tiêu chí lựa chọn tài liệu số hóa khi được ban hành có thể đối phó một cách linh hoạt với những thay đổi trong công tác quản lý, nhân sự, ví dụ các chính sách của đơn vị thay đổi thì với những tiêu chí đã được xây dựng, có thể dễ dàng điều chỉnh, bổ sung, thay thế.. những tiêu chí bị ảnh hưởng, hoặc trong quá trình thực hiện công việc hoàn toàn có thể điều chỉnh nhân lực thực hiện mà không quá phụ thuộc vào một cá nhân hay một nhóm nào.

Ngoài ra, với văn bản này sẽ giúp nhóm/bộ phận thực hiện công việc một cách hiệu quả hơn, thống nhất hơn bởi tất cả các thành viên sẽ bắt buộc phải thực hiện theo đúng quy định. Và như vậy văn bản này sẽ là một “*tài liệu mẫu*” cho các dự án số hóa tiếp theo, sẽ đảm bảo tính thống nhất cho tất cả các dự án số hóa, hoặc xây dựng các bộ sưu tập số.

### ***2.3. Giải quyết hiệu quả vấn đề bản quyền***

Hiện nay dịch vụ số hóa toàn văn tài liệu là một trong số những dịch vụ đang rất phổ biến trong các cơ quan thông tin thư viện ở nước ta. Tuy nhiên, mọi người, kể cả người cung cấp dịch vụ lẫn người sử dụng dịch vụ đều chưa quan tâm nhiều đến khía cạnh bản quyền trong quá trình triển khai dịch vụ mà đặc trưng của tài liệu ở dạng số là dễ sao chép, dễ phổ biến, dễ lưu giữ.

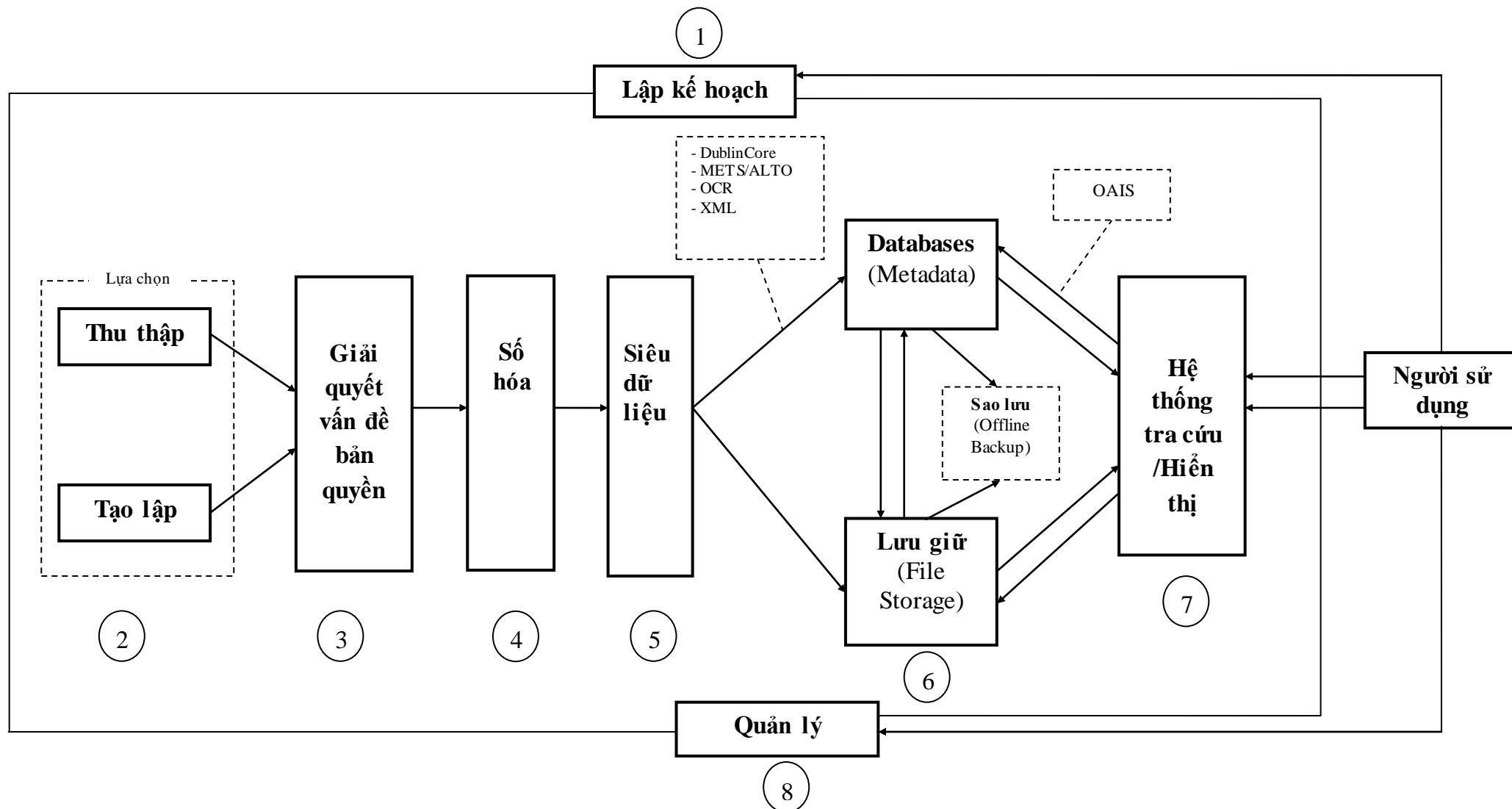
Đây là vấn đề cần quan tâm và giải quyết, bởi việc số hóa tài liệu với bất cứ mục đích nào cũng ảnh hưởng nhất định đến quyền lợi của của các chủ thể liên quan, đó là thực tế, vì lẽ đó mà ngay sau mỗi trang tên sách, vấn đề bản quyền luôn được

khuyến cáo vi phạm, trừ những cuốn sách đã hết thời gian bảo hộ bản quyền hoặc những tài liệu không xác định được bản quyền tác giả.

Vấn đề bản quyền trong việc số hóa tài liệu, cũng như bản quyền trong môi trường kỹ thuật số nói chung rất phức tạp. Để hiểu rõ và thực hiện đúng vấn đề bản quyền, cần làm rõ các vấn đề sau:

- Đối tượng bảo hộ quyền tác giả và quyền liên quan
- Tác giả, chủ sở hữu quyền tác giả, quyền liên quan
- Xin cấp phép sử dụng quyền tác giả và quyền liên quan
- Giới hạn và ngoại lệ

Để giải quyết các vấn đề pháp lý khác phát sinh khi số hóa tài liệu hoặc phổ biến tài liệu đã số hóa, các tư viện cần thành lập một ban nghiên cứu về bản quyền tài liệu, ngoài công tác nghiên cứu, nắm bắt các quy định cụ thể về bản quyền tác giả và sở hữu trí tuệ, ban này còn tư vấn cho lãnh đạo, hoặc là đầu mối giải quyết các vấn đề pháp lý phát sinh trong công tác xây dựng các bộ sưu tập số.



Mô hình giải pháp xây dựng bộ sưu tập số hiệu quả

## ***2.4. Quy chuẩn công tác số hóa tài liệu***

Có thể nói, số hóa tài liệu là một khâu đơn giản nhất nhưng thú vị nhất của một dự án số hóa tài liệu. Thú vị ở đây là người thực hiện số hóa có thể được tiếp xúc trực tiếp với những loại tài liệu quý hiếm và có thể nhìn thấy “cuộc sống” của tài liệu đó trên màn hình máy tính sau khi đã số hóa. Đơn giản ở đây có thể hiểu là nếu bạn muốn quét (scan) hoặc một trang tài liệu ở mức độ căn bản thì khá dễ dàng, tuy nhiên nếu dự án yêu cầu số hóa tài liệu với hình ảnh chất lượng cao thì vấn đề lại trở nên khá phức tạp.

Số hóa tài liệu cũng là công việc lặp đi lặp lại dễ gây nhàm chán, có thể gây ra những sai lầm nghiêm trọng, và việc khắc phục sẽ gây tốn kém hoặc mất nhiều thời gian do đó chất lượng hình ảnh số hóa cần phải được quy định ngay từ đầu.

Để hạn chế sai lầm trong công tác số hóa tài liệu, cần thiết phải áp dụng triết lý “Số hóa một lần cho tất cả” thống nhất tất cả các dự án hoặc các công việc nhỏ, lẻ khác trong kế hoạch hoặc không trong kế hoạch đều phải đảm bảo hai mục đích chính “Số hóa để bảo quản” và “Số hóa để phục vụ”, khi đó hình ảnh được số hóa phải được thực hiện ở chất lượng cao nhất có thể (mặc dù biết rằng chất lượng của sản phẩm còn phụ thuộc vào điều kiện hạ tầng, trang thiết bị số hóa...), bởi chất lượng số hóa càng cao, đồng nghĩa với việc tuổi thọ càng dài và sản phẩm có thể được sử dụng cho nhiều mục đích khác nhau. Vì vậy cách tốt nhất là chỉ đơn giản là "Số hóa một lần" để tạo ra tệp chủ (master file) và tạo những tệp phái sinh (copy file) từ những hình ảnh gốc cho những mục đích khác tùy theo từng công việc cụ thể.

### ***2.4.1. Số hóa hình ảnh***

Việc tạo hình ảnh màu đã trở thành chuẩn mực đối với các máy ảnh và máy scan kỹ thuật số trong thời đại ngày nay, và chi phí cho mỗi MB dữ liệu lưu trữ đã giảm đáng kể so với trước đây. Có thể giải quyết vấn đề chất lượng hình ảnh bằng cách sử dụng các giá trị mục tiêu hoặc các cơ chế kiểm soát chuẩn hóa khi tạo hình ảnh. Hiện tại, việc lưu trữ các hình ảnh nặng 48-bit (không nên nhầm lẫn với kỹ thuật số hóa 48-bit) chỉ phù hợp với một số ít trường hợp. Chất lượng hình ảnh ở mức độ này chỉ có ưu điểm khi muốn tạo ra các bản scan có chất lượng cực kỳ cao, mà điều này thì không phù hợp và không cần thiết đối với phần lớn các loại tư liệu, tài liệu.

### ***2.4.2. Các thông số kỹ thuật trong số hóa***

Việc số hóa nhằm mục đích sao chép một bản in hoặc một văn tự nguyên bản với chất lượng càng xác thực càng tốt và phù hợp với các yêu cầu khoa học áp dụng trong lĩnh vực này. Cần lựa chọn các thông số cho quá trình số hóa liên quan đến chất

lượng hình ảnh, khả năng lưu trữ và sử dụng trong dài hạn, và khả năng tương thích với công nghệ hiện tại và tương lai.

Trong kỹ thuật số hóa, có hai hình thức sao chép tư liệu quan trọng. Thứ nhất là sao chép các tệp chủ (master file) dưới dạng số, tức là ở định dạng thô hoặc định dạng lưu trữ. Thứ hai là những tệp phái sinh (copy file) được tạo ra cho người sử dụng, vốn là những bản copy đã được giảm dung lượng và được đưa về các dạng file khác. Các thông số dưới đây chỉ áp dụng cho các tệp chủ số. Nên tạo ra các tệp phái sinh như các file dạng JPEG hoặc GIF từ tệp chủ tùy thuộc vào mục đích sử dụng. Giả sử các tệp chủ có chất lượng thỏa đáng, thì các chế bản hoàn toàn có thể được tùy chỉnh theo nhu cầu. Điều này có thể cần thiết, ví dụ như trong trường hợp độ phân giải màn hình phía bên người sử dụng có sự thay đổi so với giả định ban đầu, hoặc khi các định dạng hình ảnh được sử dụng với các đặc tính được tối ưu hóa nhằm đáp ứng một mục đích trình bày nhất định (phóng to - zoom) liên tục, chuyển tiếp tự nhiên giữa các mảng của các đối tượng ảnh có kích thước lớn và nhiều chi tiết như bản đồ hoặc các văn bản cổ xưa.

Tệp chủ chính là nền tảng cho tất cả các quy trình xử lý về sau. Do đó, cần đặc biệt chú ý đến quá trình tạo và lưu trữ tệp chủ. Tuy nhiên, ngay cả đối với các tệp chủ cũng không nên mặc định phải đạt chất lượng kỹ thuật tốt nhất có thể ở thời điểm tiến hành.

#### 2.4.2.1. Độ phân giải và chất lượng hình ảnh

Đối với hình màu hoặc tông xám, nên sử dụng độ phân giải tương ứng với định dạng của tư liệu gốc nhưng không nhỏ hơn 300 dpi. Đối với các bản văn tự nguyên bản hoặc bản đồ có chứa những đường nét hoặc chữ viết rất mảnh, cần sử dụng độ phân giải 400 dpi. Các bản scan nhị sắc cần có độ phân giải 600 dpi.

Các mức phân giải cao hơn thường hiếm khi có tác dụng bởi vì các giá trị tiêu chuẩn nêu trên thường đã đủ để đảm bảo khả năng hiển thị toàn bộ các thông tin quan trọng. Đối với những trường hợp thẩm định đặc biệt, ví dụ khi thẩm định cấu trúc giấy đòi hỏi phải phóng đại ở mức độ lớn, thì có thể phải áp dụng độ phân giải cao hơn; tuy nhiên những trường hợp như vậy nằm ngoài phạm vi các khuyến nghị này.

Mặc dù vậy, độ phân giải cũng chỉ là một trong một số khía cạnh quyết định chất lượng hình ảnh. Ngoài ra còn một khía cạnh khác – đó chính là yếu tố kỹ thuật. Do đó, cần kiểm tra kỹ hình ảnh được tạo ra để đảm bảo sự trung thực về màu sắc và trung thành với nguyên gốc. Để phục vụ mục đích này, khuyến nghị nên hiệu chuẩn màn hình và thiết lập một môi trường ánh sáng có kiểm soát, để có thể đánh giá hình ảnh trên màn hình một cách khách quan.



Khi sử dụng máy scan, khoảng cách đến đối tượng đích đã được cố định. Độ phân giải của máy scan luôn không đổi trong giới hạn kích thước tối đa đã được mặc định cho đối tượng (ví dụ độ phân giải 300 dpi có thể được áp dụng cho kích thước tối đa A3). Đối với máy ảnh kỹ thuật số, độ phân giải phụ thuộc vào khoảng cách đến đối tượng đích. Để đảm bảo độ phân giải 300 dpi, cần tính toán một lần kích thước tối đa của đối tượng cho một máy ảnh nhất định. Ví dụ, nếu máy ảnh có ma trận 4.000 × 3.000 điểm ảnh (tương đương 12 triệu điểm ảnh) thì chỉ có thể chụp được những đối tượng có kích thước nhỏ hơn hoặc bằng 33,9 x 25,4 cm. Có thể tính toán theo công thức:

$$\frac{\text{Số lượng điểm ảnh (điểm)}}{\text{Độ phân giải (dpi)}} = \text{kích thước tối đa của đối tượng (đơn vị inch, 1 inch = 2,54 cm)}$$

*Đối với độ phân giải 300 dpi:*

Chiều dài: 4.000 điểm ÷ 300 dpi = 13,3 inch = 33,9 cm

Chiều rộng: 3.000 điểm ÷ 300 dpi = 10 inch = 25,4 cm

Khi số hóa phim, cần cân nhắc hệ số giảm mòn của vi phim (microfilm) hoặc vi phích (microfiche) tương ứng với bản gốc để đảm bảo đạt được độ phân giải đích là 300 dpi. Ví dụ, nếu bản gốc có kích thước 24 x 36 cm, và được ghi lại trên phim dương bản truyền thống cỡ 24 × 36 mm, thì phải nhân độ phân giải đích với hệ số 10 khi scan từ phim; nói cách khác, phim dương bản cần phải được scan ở độ phân giải 3.000 dpi. Để xác định chính xác độ phân giải cần scan, cần biết rõ kích thước của đối tượng gốc hoặc ít nhất cần ước tính một cách khoa học và hợp lý kích thước này (ví dụ: kích thước trang không quá 40 cm) để tránh tình trạng sao chép không đảm bảo được phân giải đích tương ứng với kích thước của bản gốc nhưng không nhỏ hơn 300 dpi.

#### 2.4.2.2. Độ sâu của màu sắc

Các bản scan nhị sắc (đen/trắng) thường sử dụng độ sâu màu sắc là 1 cấp (1 bit) trên một điểm ảnh. Do đó, mỗi điểm ảnh sẽ tồn tại dưới dạng 1 (= đen) hoặc 0 (= trắng). Hình ảnh có tông màu xám được số hóa ở độ sâu màu sắc 256 cấp trên một điểm ảnh. Hình ảnh màu sử dụng ba kênh màu đỏ, xanh lục và xanh lam (trong kỹ thuật gọi là màu RGB), do đó sẽ có độ sâu màu sắc kết hợp bằng 3 lần 256 cấp. Vì thế, xét từ góc độ kỹ thuật một hình ảnh màu lớn gấp 3 lần một hình ảnh tông xám, trong đó các giá trị màu cho mỗi kênh được cộng lại với nhau và được áp dụng cho mỗi điểm ảnh (ví dụ 35 đỏ + 233 xanh lục + 186 xanh lam), cho kết quả tổng cộng là 16,7

triệu màu ( $256 \times 256 \times 256$ ). Sự khác biệt 256 cấp đòi hỏi 8 bit, tương đương với 1 byte, trong máy tính (mỗi bit tồn tại dưới dạng 0 hoặc 1). Do đó, một hình ảnh màu sẽ có độ sâu màu thông thường là 24 bit ( $3 \times 8 \text{ bit} = 3 \text{ byte}$ ). Một số nhà sản xuất máy ảnh và máy scan có thể cho ra những sản phẩm có độ sâu màu sắc cao hơn, lên đến 48-bit. Mức độ này thực sự không cần thiết bởi vì độ sâu màu sắc 24-bit đã là hoàn toàn đủ để thể hiện trên màn hình máy tính, tuy nhiên độ sâu 48-bit cũng có thể hữu dụng khi biên tập một hình ảnh scan, bởi lẽ các giá trị màu có thể bị mất đi trong quá trình chỉnh sửa cấp độ. Chụp hình ở 48 bit thường có ưu điểm hơn và cho hình ảnh cân đối hơn, bởi máy ảnh nhận biết phổ màu khác với mắt người.

Tuy nhiên, lưu trữ ảnh 48-bit lại hiếm khi là một việc làm cần thiết. Cần cân nhắc thêm rằng một hình ảnh 48-bit sẽ chiếm gấp đôi không gian lưu trữ, và điều này có thể trở thành một vấn đề thực sự nan giải đối với những bộ tài liệu lưu trữ khối lượng lớn. Do đó, nên hạn chế những hình ảnh có độ sâu màu sắc trên 24 bit và chỉ áp dụng độ sâu này cho những tư liệu đòi hỏi chất lượng màu chính xác nhất do những lý do khoa học, những không gian màu đặc thù, hoặc để phục vụ công tác biên tập toàn diện có thể sẽ diễn ra.

#### 2.4.3. Định dạng tệp

Theo công nghệ hiện tại, các tệp chủ hình ảnh với tông màu xám hoặc hình ảnh màu nên được lưu trữ dưới dạng file TIFF (Không nén). Đối với những hình ảnh nhị sắc, có thể sử dụng dạng file TIFF với đặc tính Nén file Nhóm 4 (TIFF Group 4). Định dạng file TIFF đã tồn tại từ những năm 1980 đến nay, và đã khẳng định vị thế của một trong những chuẩn định dạng quan trọng nhất. Dự kiến toàn bộ các chương trình máy tính chuẩn sẽ tiếp tục hỗ trợ định dạng file này. Tuy nhiên, điều này chỉ đúng đối với những file có định dạng TIFF cơ bản (Baseline TIFFs). Những thuật ngữ này xuất phát từ một thực tế: định dạng TIFF chuẩn nói chung là một định dạng file cực mạnh, cho phép lưu cả những hình ảnh có đặc tính khác biệt – ví dụ như những hình ảnh được phân chia thành các lát (tile) để có thể tải từng phần của bức hình một cách độc lập, vốn là một đặc tính cực kỳ hữu dụng cho các lĩnh vực như tổng hợp hình ảnh bản đồ đòi hỏi độ phân giải cao. Tuy nhiên, một định dạng file mạnh như vậy rất khó triển khai toàn phần. Vì lý do này, định dạng chuẩn TIFF có sự phân biệt giữa một nhóm nhỏ các thuộc tính hình ảnh cần phải được hỗ trợ bởi những phần mềm ứng dụng được cho là có thể hỗ trợ TIFF với một số lượng mở rộng các thuộc tính mà phần mềm ứng dụng nào cũng có thể sử dụng.

Nếu khó khăn nằm ở không gian lưu trữ, thì về mặt lý thuyết, hình ảnh có thể được lưu trữ bằng phương pháp nén mà không sợ dữ liệu bị mất đi (ví dụ định dạng

TIFF nén LZW – LZW-compressed TIFF). Tuy nhiên, định dạng nén bao giờ cũng đi kèm một rủi ro nhất định, bởi vì chỉ cần một hư hại nhỏ nhất xảy ra với file hoặc khi một vài bit bị đảo cũng có thể gây ảnh hưởng đến chất lượng của toàn bộ hình ảnh. Những hư hại này có thể xảy ra do phương tiện lưu trữ bị hỏng, hoặc khi sao chép file. Điều này đặc biệt phổ biến với định dạng file JPEG. Chúng ta thường gặp những trường hợp file JPEG bị hỏng và chỉ có thể hiện thị một phần hình ảnh. Những hình ảnh bị hư hại như thế này thường không thể khắc phục được. Cái khó ở đây là hình ảnh đó có thể hoàn toàn ổn khi nằm trên một phương tiện, nhưng lại bị hư hại trong quá trình sao chép và chuyển sang một phương tiện lưu trữ mới. Do đó, JPEG không phải là một định dạng phù hợp cho việc lưu trữ. Nếu vì lý do chi phí mà không thể tránh khỏi yêu cầu nén file, thì nên sử dụng định dạng file PNG hoặc TIFF (LZW).

Một bước phát triển mới là JPEG2000, không chỉ cho phép thực hiện thuật toán nén file theo phương pháp mới và hiệu quả hơn, mà kết quả cũng chắc chắn hơn so với JPEG. Không giống với JPEG thông thường, JPEG2000 cho phép lưu trữ mà hoàn toàn không lo bị mất dữ liệu. Ngoài ra, JPEG2000 còn cho phép truyền hình ảnh với tốc độ nhanh chóng (hình tải được chừng nào, thì cũng có thể thấy luôn chừng đó chi tiết), và có thể chứa cả lý lịch dữ liệu. Nhưng có lẽ, lợi điểm ưu việt nhất của JPEG2000 chính là khả năng từ một hình ảnh lớn tạo ra thành rất nhiều độ phân giải và thậm chí cả những chi tiết khác nhau, nhờ đó công tác quản lý việc lưu trữ hình ảnh trở nên đơn giản hơn bởi không còn phải lưu trữ các hình ảnh ở các độ phân giải khác nhau nữa.

Để phục vụ việc ấn hành hoặc xuất bản trên Internet, JPEG và PNG được khuyến dùng do phổ biến và được nhiều người sử dụng ưa chuộng. GIF, do bảng màu hạn chế, nên chỉ phù hợp khi scan nhị sắc hoặc tông xám. Nhưng suy cho cùng thì muốn sử dụng định dạng gì cũng đều được bởi tất cả các định dạng được hỗ trợ bởi các trình duyệt chuẩn đều có thể được tạo ra một cách đơn giản từ tệp chủ. Do đó, nếu một quyết định sai lầm chẳng may được đưa ra thì cũng có thể điều chỉnh dễ dàng, và khi điều kiện thay đổi cũng có thể nhanh chóng ứng phó.

### ***2.5. Lựa chọn siêu dữ liệu phù hợp***

Siêu dữ liệu (metadata) dùng để mô tả một tài nguyên thông tin được chia sẻ trên Internet. Một bản ghi metadata bao gồm một tập các thuộc tính hoặc tập các phần tử cần thiết để mô tả các tài nguyên theo yêu cầu. Ví dụ, một hệ thống metadata thông thường trong biên mục thư viện chứa một tập các bản ghi metadata dùng để mô tả sách như : tác giả, nhan đề, thời gian xuất bản, chủ đề-đề mục, chỉ số xác định vị trí trên kho, giá v.v....

Có nhiều chuẩn mô tả biên mục mang tính chất metadata khá thông dụng, thí dụ MARC21/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 (search engine) tìm kiếm, lọc ra các thông tin metadata để tổ chức thành các kho dữ liệu mà không cần dùng đến hệ quản trị cơ sở dữ liệu truyền thống. Nói cách khác, ngay bản thân ngôn ngữ XML tự nó đã hỗ trợ cho việc hình thành một cơ sở dữ liệu toàn văn bản, phi cấu trúc rất tiện lợi cho việc tìm kiếm và trao đổi thông tin.

\* Đối với tài liệu số hóa dạng sách sử dụng chuẩn DublinCore để mô tả tài liệu.

Dublin Core Metadata là chuẩn dùng mô tả nội dung của biểu ghi và dữ liệu. Nó đơn giản hơn MARC Format rất nhiều vì chỉ có 15 phần tử, trong khi MARC có đến hơn 200 trường. Mục đích của việc biên mục siêu dữ liệu là để sử dụng trên mạng Internet. Các phần tử dữ liệu MARC và Dublin Core có thể trao đổi lẫn nhau theo một giản đồ quy định khi hiển thị cho người sử dụng. Các phần tử này là:

1. Nhan đề (Title): tên gọi chính thức của tài liệu
  2. Tác giả (Creator): tên của một hay một số tác giả chính
  3. Chủ đề (Subject): tên chủ đề - đề mục dùng để phân loại tài liệu
  4. Mô tả (Description): mô tả vắn tắt nội dung tài liệu
  5. Nhà xuất bản (Publisher): tên gọi, nơi ban hành tài liệu
  6. Tác giả phụ (Contributor): tên của những tác giả cộng tác
  7. Ngày tháng (Date): ngày tháng ban hành tài liệu
  8. Kiểu tài liệu (Type): kiểu tài liệu
  9. Mô tả vật lý (Format) : Thông tin về dạng vật lý
  10. Định danh (Identifier): Thông tin định danh của tài liệu
  11. Nguồn gốc (Source): Thông tin về xuất xứ của tài liệu
  12. Ngôn ngữ (Language): Thông tin về ngôn ngữ
  13. Liên kết (Relation): Thông tin liên kết của tài liệu
  14. Bao quát (Coverage): Thông tin liên quan diện bao quát
  15. Bản quyền (Rights): Thông tin liên quan quyền tác giả
- Các yếu tố siêu dữ liệu Dublin Core có những ưu điểm sau:

+ *Tạo lập và sử dụng dễ dàng*: cho phép những người không chuyên nghiệp có thể tạo các bản ghi mô tả đơn giản cho các tài nguyên thông tin và truy xuất chúng trên môi trường mạng một cách dễ dàng.

+ *Ngữ nghĩa dễ hiểu, sử dụng đơn giản*: Việc khai thác thông tin trên mạng Internet diện rộng thường gặp trở ngại bởi những sự khác nhau về thuật ngữ và sự mô tả thực tế. Dublin Core Metadata giúp những người dò tìm thông tin không chuyên có thể tìm thấy vấn đề mình quan tâm bằng cách hỗ trợ một tập hợp các phần tử thông dụng mà ngữ nghĩa của chúng được hiểu phổ biến.

+ *Phạm vi phổ biến*: Tập hợp các phần tử Dublin Core Metadata lúc đầu được phát triển bằng tiếng Anh, nhưng hiện nay nó được cập nhật thêm với khoảng 25 ngôn ngữ khác nhau (phiên bản v1.1)

+ *Tính mở rộng*: Những nhà phát triển Dublin Core đã cung cấp một cơ chế cho việc mở rộng tập các phần tử Dublin Core, phục vụ nhu cầu khai thác các tài nguyên bổ sung. Các phần tử Metadata từ những tập các phần tử khác nhau có thể liên kết với metadata của Dublin Core. Điều này cho phép các tổ chức khác nhau có thể dùng các phần tử Dublin Core để mô tả thông tin thích hợp cho việc sử dụng tài nguyên trên Internet.

+ *Giúp nâng cao độ chính xác của định chỉ số*

+ *Có khả năng liên tác (Interoperability), sử dụng lẫn nhau*

+ *Mở rộng thuận lợi*

\* Đối với tài liệu số hóa dạng báo, tạp chí, sử dụng chuẩn METS/ALTO để mô tả.

**METS** (Tiêu chuẩn truyền và Mã hóa siêu dữ liệu - Metadata Encoding and Transmission Standard).

Một tiêu chuẩn METS bao gồm 5 phần chính:

1. Siêu dữ liệu mô tả: phần siêu dữ liệu mô tả có thể chỉ ra siêu dữ liệu mô tả bên ngoài, hoặc bao gồm siêu dữ liệu mô tả mã hóa bên trong hoặc cả hai.

2. Siêu dữ liệu quản trị: phần siêu dữ liệu quản trị cung cấp thông tin về tạo dựng và lưu giữ các tệp như thế nào, quyền sở hữu trí tuệ, siêu dữ liệu về nguồn gốc của đối tượng thư viện số và thông tin về các tệp gốc cấu tạo nên đối tượng thư viện số (Ví dụ: Mối quan hệ giữa tệp gốc và tệp phái sinh, thông tin về chuyển đổi và thay thế). Giống siêu dữ liệu mô tả, siêu dữ liệu quản trị có thể chỉ ra cả siêu dữ liệu quản trị bên ngoài lẫn siêu dữ liệu mã hóa bên trong.

3. Nhóm tệp: phần nhóm tệp liệt kê tất cả các tệp cấu tạo nên các phiên bản điện tử của đối tượng số. Các phần tử của nhóm tệp giúp tham chiếu đến các tệp cấu tạo của phiên bản số.

4. Sơ đồ cấu trúc: Sơ đồ cấu trúc là trung tâm chính của một tệp METS. Phần này mô tả cấu trúc thứ bậc của đối tượng thư viện số, tạo liên kết giữa các phần tử của cấu trúc đó với các tệp nội dung và siêu dữ liệu liên quan.

5. Cơ chế vận hành: Bộ phận vận hành được sử dụng để liên kết các hoạt động quản lý với phần nội dung của một đối tượng được mô tả theo tiêu chuẩn METS. Phần này có 1 bộ phận xác định giao diện có thể mô tả cụ thể, chính xác các quy trình hoạt động. Ngoài ra nó còn có một cơ chế vận hành, đó là một bộ phận gồm các mã vận hành có nhiệm vụ duy trì và thực hiện các hoạt động mà bộ phận xác định giao diện đã lập ra.

METS được sử dụng vào những mục đích chính sau:

- Tạo tài liệu XML chứa đựng cấu trúc hình cây của những đối tượng tài liệu số.
- Ghi lại tên và nơi lưu giữ các tập tin tạo thành những đối tượng tài liệu số đó.
- Ghi lại những metadata liên quan để có thể sử dụng như một công cụ mô hình hóa những đối tượng trong thế giới vật lý, như những loại tài liệu cụ thể.

Tiêu chuẩn METS cung cấp một cơ chế linh hoạt để mã hóa siêu dữ liệu mô tả, quản trị và cấu trúc cho một đối tượng số trong thư viện, và để thể hiện các mối quan hệ phức tạp giữa các hình thức khác nhau của siêu dữ liệu. Bởi vậy nó có thể cung cấp 1 tiêu chuẩn thống nhất cho việc trao đổi các đối tượng số giữa các kho trong thư viện.

Những thư viện học thuật và nghiên cứu hàng đầu hiện nay đang trích dẫn METS như là một tiêu chuẩn quan trọng để vận hành gắn kết lẫn nhau trong một thư viện số, và ngày nay nó đang được cộng đồng thư viện số trên thế giới sử dụng rộng rãi.

### **ALTO** (Analyzed Layout and Text Object)

Trong khi tiêu chuẩn METS được sử dụng để mô tả cấu trúc của một tài liệu số thì ALTO được sử dụng để cất giữ thông tin dàn trang và các ký tự được nhận máy tính nhận dạng (OCR) trong các trang của những tài liệu in như sách, báo và tạp chí. Nó được thiết kế để được sử dụng như một lược đồ mở rộng cho lược đồ XML Metadata Encoding and Transmission Schema (METS) của thư viện Quốc hội Mỹ. Trong đó METS cung cấp các thông tin cấu trúc còn ALTO chứa đựng nội dung và các thông tin vật lý của tài liệu.

Chuẩn ALTO đã mang đến rất nhiều lợi ích cho việc lưu giữ các tài liệu số như:

- Việc tìm kiếm toàn văn trở nên chính xác hơn khi việc tìm kiếm chỉ được thực hiện trong phần nội dung cuốn sách mà không bị “nhiều” vì những thông tin như: trang bìa tài liệu, các tiêu đề cột, chú thích v.v...

- Việc tìm kiếm theo cấu trúc có thể được thực hiện như tìm kiếm trong các chương hoặc đối với tạp chí là trong những loạt bài viết.

- Việc định hướng trong cuốn sách trở nên dễ dàng hơn.

- Các hình minh họa, các công thức, bảng biểu và thậm chí cả các quảng cáo cũng trở thành những thành phần riêng và có thể được truy cập riêng rẽ với phần nội dung còn lại.

## **2.6. Xây dựng kế hoạch bảo quản số**

Mục tiêu của bảo quản số là đảm bảo nó giữ nguyên khả năng truy cập cho người sử dụng, do đó khi xây dựng quy trình bảo quản tài liệu số, cần tuân thủ chặt chẽ những vấn đề sau:

+ Cần có chiến lược và chính sách bảo quản: Chiến lược và chính sách bảo quản số cần phải được phát triển, có tính tới mức độ khẩn cấp, hoàn cảnh địa phương, các phương pháp hiện có và kế hoạch tương lai. Sự hợp tác của người giữ quyền tác giả và các quyền có liên quan, và người đại diện khác, trong việc thiết lập tiêu chuẩn chung và tính tương hợp, và chia sẻ nguồn lực sẽ tạo điều kiện thuận lợi cho việc này.

+ Lưu giữ có lựa chọn: Như với tất cả di sản tư liệu, nguyên tắc lựa chọn có thể thay đổi giữa các quốc gia, mặc dù tiêu chuẩn chính để quyết định tài liệu nào cần lưu giữ phải là tầm quan trọng và việc chứa đựng giá trị văn hóa, khoa học, bằng chứng và giá trị khác. Tài liệu “kỹ thuật số sinh ra – born digital” rõ ràng phải được ưu tiên. Quyết định lựa chọn và xem xét tiếp sau bất kỳ cần được tiến hành có trách nhiệm, và cần dựa trên các nguyên tắc, chính sách, quy trình và tiêu chuẩn xác định.

+ Cần bảo vệ và bảo quản số: cần có bộ khung pháp lý và thể chế thích hợp để bảo vệ di sản số như là một thành phần chủ yếu của chính sách bảo quản quốc gia.

Hiện tại, để bảo quản lâu dài các nội dung số hóa, không có một giải pháp nào phù hợp cho tất cả các loại hình đối tượng và tư liệu khác nhau. Các tiêu chí then chốt để bảo quản thành công tài liệu số trong dài hạn bao gồm:

(1) Khả năng tạo ra một khuôn khổ tổ chức và kinh tế phù hợp.

(2) Khả năng thiết lập một môi trường kỹ thuật phù hợp, kết hợp lựa chọn các kỹ thuật và chiến lược phù hợp.

Mặt khác, kết quả của các dự án số hóa có đảm bảo độ an toàn trong dài hạn hay không còn phụ thuộc vào sự lựa chọn dữ liệu và định dạng lý lịch dữ liệu.

Vấn đề lưu trữ thường không được quan tâm đúng mực, và chi phí cũng như công sức cần thiết cho công tác này thường bị ước tính thấp hơn nhu cầu thực tế. Máy chụp ảnh hoặc máy scan có độ phân giải càng cao thì càng tốn nhiều không gian để lưu trữ các hình ảnh được tạo ra. Tập chủ (master file) dưới dạng kỹ thuật số của một hình ảnh màu có thể nặng từ 20-80 megabyte (MB) hoặc hơn thế. Những tư liệu hình ảnh nặng vài terabyte (TB) có thể nhanh chóng đầy lên và cần phải được lưu trữ.

Hiện tại, có bốn phương tiện phổ biến được sử dụng cho hầu hết các hình thức lưu trữ dữ liệu kỹ thuật số:

- (1) Phương tiện quang học có thể tháo lắp được, như CD và DVD
- (2) Ổ băng (ổ băng chạy suốt)
- (3) Ổ cứng
- (4) Vi phim (Microfilm)

Đối với trường hợp (1): Lưu trữ trên các phương tiện quang học có thể tháo lắp được như CD-R hoặc DVD-R không hiệu quả đối với các dự án số hóa quy mô lớn. Bên cạnh đó, CD và DVD cũng không phải là các phương tiện phù hợp để bảo quản dữ liệu trong dài hạn, khi mà công nghệ luôn thay đổi nhanh chóng và tư liệu có sự nhất quán. Lưu trữ bằng CD hoặc DVD có thể phù hợp đối với những người mới bắt đầu xây dựng một kho lưu trữ kỹ thuật số hoặc chỉ muốn số hóa một số tư liệu hữu hạn. Nhưng khi khối lượng dữ liệu tích lũy nhiều hơn, thì sẽ rất nhanh chóng dùng hết giới hạn dung lượng. Đĩa CD chuẩn có thể chứa 700 MB, DVD có thể chứa 4,7 GB hoặc xấp xỉ 8 GB nếu sử dụng kỹ thuật hai lớp. Chỉ cần lưu trữ 1 terabyte (tương đương với khoảng 180 quyển sách) cũng đã cần đến 1.429 đĩa CD, 213 đĩa DVD dung lượng 4,7 GB, hoặc 125 đĩa DVD dung lượng 8 GB. Nếu nhìn vào những con số này, có thể thấy việc lưu trữ bằng CD sẽ nhanh chóng trở nên cồng kềnh. Nhưng ngay cả DVD cũng không phải là một giải pháp thực sự thỏa đáng để lưu trữ dữ liệu với khối lượng lớn. Cần ghi đĩa (burn), dán nhãn, và lưu trữ, và những nhiệm vụ này có thể tốn rất nhiều thời gian. Khả năng truy cập trực tiếp cũng trở nên phức tạp và phụ thuộc vào địa điểm, trừ khi sử dụng các robot hoặc máy chủ CD (CD server). Nếu sử dụng CD/DVD để lưu trữ, cần định kỳ tiến hành kiểm tra đột xuất và ngẫu nhiên, đặc biệt khuyến nghị nên lưu trữ thừa ra (ít nhất hai bản).

Đối với trường hợp (2): Lưu trữ bằng băng là phương pháp có phần tiện lợi hơn khi lưu trữ đại trà với số lượng lớn; tuy nhiên, phương pháp này cũng có một nhược



điểm là băng có tốc độ tương đối chậm. Trong trường hợp cần thường xuyên truy cập bản master thì băng không phải là một giải pháp tốt. Hơn nữa, phải thường xuyên chạy băng để tránh bị dính băng. Các thư viện của các trường đại học – hoặc thư viện của những tổ chức liên quan/ liên kết với các trường đại học – nếu muốn triển khai hệ thống lưu trữ bằng băng thì nên tham khảo ý kiến từ trung tâm máy tính của nhà trường trước khi quyết định chiến lược lưu trữ dài hạn. Các hệ thống lưu trữ băng hiện đại (sử dụng robot) được thiết kế để có thể lưu trữ vài trăm terabyte và có khả năng đảm bảo sao chép mỗi băng thành một vài bản cũng như chạy băng với tần suất cần thiết.

Đối với trường hợp (3): Khi xem xét khả năng di chuyển, nên lưu trữ dư thừa trên đĩa cứng (ví dụ RAID 5) dưới dạng hệ thống Lưu trữ qua Mạng (Network Attached Storage - NAS) hoặc Mạng Lưu trữ (Storage Attached Networks - SAN) tại các trung tâm dữ liệu. Vì lý do an toàn, cần tạo băng dự phòng (tape backup) hoặc tạo thêm một bản sao đĩa cứng (disk copy). Băng dự phòng và bản sao đĩa cứng cần được cất ở cùng một nơi. Mô hình này cho phép truy cập dữ liệu nhanh chóng, đơn giản và tạo thuận tiện mỗi khi cần di chuyển. Mô hình lưu trữ bằng ổ cứng dựa trên giả định dữ liệu được cập nhật liên tục, và không có dữ liệu nào bị gạt sang một bên mà không có những quy trình xử lý phần cứng hoặc phần mềm đảm bảo tính toàn vẹn. Xét ở khía cạnh đó, mô hình này có cách tiếp cận khác nhưng lại chưa được kiểm chứng đối với việc lưu trữ dài hạn các phương tiện số - mà suy cho cùng thì đây vẫn là một vấn đề chưa ngã ngũ.

Đối với trường hợp (4): Một phương pháp khác có thể sử dụng để lưu trữ bản sao các tư liệu số trong dài hạn là in lên vi phim (microfilm) và sử dụng chiến lược chuyển đổi. So với các chất liệu ảnh tương tự khác, vi phim có độ bền cao nhất khi lưu trữ trong điều kiện tối ưu. Nếu cần, sau này có thể chuyển các vi phim này ngược trở lại dạng kỹ thuật số với sự hỗ trợ của thiết bị scan phim. Ngoài ra, hiện còn có những giải pháp cho phép in các bản sao màu kỹ thuật số với chất lượng cao lên vi phim màu. Tuy nhiên, hiện chưa thể đánh giá và đưa ra kết luận chắc chắn về những công nghệ mới phát triển này.

### ***2.7. Đẩy mạnh công tác phổ biến rộng rãi các bộ sưu tập số***

Đẩy mạnh công tác phổ biến rộng rãi các bộ sưu tập số đã có, cần thực hiện:

- Xác định chính xác được nhóm người sử dụng đối với mỗi bộ sưu tập.
- Lưu ý đến những đối tượng người sử dụng yếu thế trong xã hội: người khuyết tật, khu vực kinh tế khó khăn, kém phát triển

- Thiết kế website và các công cụ tìm kiếm, truy cập, hiển thị một cách khoa học, đảm bảo dễ dàng sử dụng.
- Chú trọng công tác tuyên truyền, quảng bá các bộ sưu tập
- Đẩy mạnh các vụ, nhất là dịch vụ cung cấp tài liệu từ xa (có thu phí hoặc miễn phí), thông qua các tài khoản người dùng và chính sách truy cập.

### **2.8. Chú trọng công tác đánh giá hiệu quả các bộ sưu tập số**

Đánh giá là quá trình đánh giá một cách có hệ thống và khách quan một dự án, chương trình hoặc một chính sách đang được thực hiện hoặc đã hoàn thành từ giai đoạn thiết kế đến triển khai và các kết quả đạt được. Mục đích của việc đánh giá là để xác định tính phù hợp và mức độ hoàn thành các mục tiêu, tính hiệu quả, tác động và tính bền vững.

Đồng thời đánh giá là công cụ quản lý quan trọng để lãnh đạo đơn vị có thể theo dõi tiến độ thực hiện và hỗ trợ quá trình ra quyết định.

Các khía cạnh có thể được đánh giá đối với một các bộ sưu tập số:

- Sự phù hợp: Các mục tiêu và mục đích xây dựng các bộ sưu tập số có phù hợp với yêu cầu không?
- Hiệu suất: Công tác tạo lập các bộ sưu tập số có được triển khai một cách kịp thời và tiết kiệm chi phí không?
- Hiệu quả: Đánh giá được mức độ đạt được các mục tiêu ở mức độ nào? Những yếu tố thuận lợi và khó khăn gặp phải trong quá trình triển khai xây dựng các bộ sưu tập số là gì?
- Tác động: Bộ sưu tập số mang lại những kết quả gì? Kết quả đó có ảnh hưởng tích cực và tiêu cực như thế nào đối với người sử dụng.
- Tính bền vững: Bộ sưu tập số có thể tồn tại lâu dài không?

Xây dựng và phát triển một các bộ sưu tập số được thực hiện qua rất nhiều công đoạn, có thể rất tốn kém và mất nhiều thời gian, do đó công tác đánh giá hiệu quả các bộ sưu tập số có vai trò hết sức quan trọng, cần thiết đối với mỗi dự án, mỗi công việc không chỉ sau khi hoàn thành dự án mà ngay trong chính quá trình thực hiện.

### **2.9. Đổi mới công tác quản lý, điều hành**

Công tác quản lý, điều hành các hoạt động liên quan đến công tác xây dựng và phát triển các bộ sưu tập số cần phải linh hoạt, ứng xử nhanh chóng đối với các tác động bên trong hoặc bên ngoài đơn vị, đồng thời cần áp dụng tối đa công nghệ thông

tin vào chỉ đạo, quản lý, điều hành giúp triển khai công việc một cách nhanh chóng, hiệu quả hơn.

Để có một dự án, một chương trình tạo lập các bộ sưu tập số có chất lượng, hiệu quả cao phụ thuộc vào sự đánh giá chính xác về nguồn nhân lực cần thiết, bởi mỗi cá nhân, mỗi bộ phận có thể mạnh, kỹ năng khác nhau, ví dụ:

+ Nhóm bảo quản, phục chế: Có nhiệm vụ đánh giá tình trạng tài liệu, có kết luận về việc tài liệu có thể tiến hành số hóa hay không, hoặc cần phải can thiệp ở những mức độ phù hợp trước khi chuyển sang bộ phận số hóa.

+ Nhóm số hóa: Có kỹ năng đánh giá được chất lượng của ảnh số sau khi scan, hoặc có những tác động, can thiệp ở mức độ phù hợp vào sản phẩm nhằm tăng cường chất lượng, đảm bảo đúng quy định

+ Nhóm xử lý, biên mục: Có kỹ năng có thể tạo ra các bản ghi siêu dữ liệu cho các tài liệu số, công tác này có ảnh hưởng trực tiếp đến việc quản lý bộ sưu tập và ảnh hưởng lớn đến việc tìm kiếm thông tin trong tài liệu.

+ Nhóm quản trị/hỗ trợ kỹ thuật: Có trình độ chuyên về công nghệ để đề xuất các giải pháp cụ thể cho từng trường hợp, có thể xây dựng chiến lược bền vững cho hoạt động vận hành thư viện số nói chung và công tác xây dựng các bộ sưu tập số nói riêng.

Như vậy trong công tác quản lý, điều hành cần có chiến lược về công tác nhân sự, đồng thời cần nhắc việc tăng thêm vai trò cho quản lý cho các nhóm các bộ phận chuyên môn để họ có thể chủ động, có trách nhiệm hơn trong mảng công việc được giao.

## **Kết luận**

Xây dựng một bộ sưu tập số hiệu quả, nhất là các dự án số hóa lớn là việc không hề dễ dàng, cần một tầm nhìn chiến lược, dựa trên xu hướng phát triển của công nghệ và các yêu cầu cấp thiết cần phải số hóa tài liệu, tuân thủ đúng các tiêu chuẩn, nguyên tắc và vận dụng một cách thống nhất, bài bản, tất cả những điều đó cần được cân nhắc kỹ, tổ chức một cách thận trọng, nghiêm túc và ngay từ đầu sẽ làm cho dự án số hóa tài liệu thành công cao nhất.

## **TÀI LIỆU THAM KHẢO**

1. Cao Minh Kiểm (2007). *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.
2. Tạ Bá Hưng (2000), *Phát triển nội dung số ở Việt Nam: Những nguyên tắc chỉ đạo*, Thông tin và Tư liệu, (1), tr. 2 – 6.
3. Chowdhury, G. G. (2003), *Introduction to Digital Libraries*, Facet Publishing, 384 pages.
4. Digital Preservation Coalition (2008), *Preservation Management of Digital Materials: The Handbook*
5. Ian H. Witte (2002), *How to Build a Digital Library*, Morgan Kaufmann Publishers, 518 pages.
6. Katherine M. Wisser (2007), *Guidelines for Digitization*, North Carolina ECHO (Exploring Cultural Heritage Online), USA
7. L. Candela (2008): *The DELOS Digital Library Reference Model - Foundations for Digital Libraries*. Version 0.98
8. M. Jordan (2006), *A Practical Guide for Libraries*, Chandos
9. National Information Standards Organization – NISO (2007), *A Framework of Guidance for Building Good Digital Collections (A NISO Recommended Practice)*, NISO, USA.
10. National Research Foundation (2010), *Managing Digital Collections: A Collaborative Initiative on the South African Framework*, National Research Foundation, South Africa.
11. Susan Schreibman (2007), *Best Practice Guidelines for Digital Collections at University of Maryland Libraries*, Office of Digital Collections and Research University of Maryland, College Park.
12. Wayne Wilson (2003), *Building and Managing a Digital Collection in a Small Library*, North Carolina Libraries, USA.