

Những kinh nghiệm từ dự án Số hóa tài liệu Pháp ngữ cổ tại Việt Nam

Nguyễn Thị Bắc

Nguyên Giám đốc Thư viện KHTH TP.HCM

Tóm tắt – Dự án số hóa kho tài liệu Pháp ngữ cổ của Thư viện Khoa học Tổng hợp Thành phố Hồ Chí Minh (Thư viện) là một bộ phận trong Dự án Thư viện điện tử Bibliotheca Vietnamica nằm trong khuôn khổ Dự án Phát huy mạng lưới phát hành sách và thư viện tại Đông Nam Á của Quỹ đoàn kết (FSP VALEASE), do Đại sứ quán Pháp tại Việt Nam và Bộ Ngoại giao Pháp phối hợp thực hiện với nhiều thư viện tại Việt Nam. Mục đích của Dự án số hóa kho tài liệu Pháp ngữ cổ của Thư viện là bảo quản di sản quốc gia, đồng thời là di sản thế giới, tránh những tác nhân gây hại và tăng cường khả năng sử dụng thông qua nội dung số. Trên trang ấn phẩm được lựa chọn từ kho sách đặc biệt của Thư viện cho Dự án này. Kinh phí do chính phủ Pháp tài trợ và công nghệ hiện đại đã được ứng dụng cho Dự án.

Bài này không đề cập đến các khía cạnh tài chính và công nghệ mà chủ yếu nhấn mạnh đến các vấn đề và mối quan tâm cần thiết khác của Dự án liên quan đến mục đích, quy mô, phạm vi và mức độ phức tạp của dự án; việc quản lý và kiểm soát dự án một cách đầy đủ trong quá trình thực hiện; và sự phối hợp chặt chẽ của Ban quản lý Dự án VALEASE, Thư viện và Công ty Dirox.

1. QUI MÔ, PHẠM VI VÀ MỨC ĐỘ PHỨC TẠP CỦA DỰ ÁN

Mục tiêu của Dự án là một sản phẩm dữ liệu số bao gồm một cơ sở dữ liệu trên 540.000 trang văn bản dạng số, một cơ sở dữ liệu hình ảnh về Việt Nam xưa lấy từ các tài liệu này, các phần mềm và công cụ tìm kiếm cần thiết.

Tài liệu cho Dự án này được chọn từ vốn tài liệu đặc biệt của thư viện.

Vốn tài liệu đặc biệt của thư viện có khoảng trên 30.000 cuốn, bao gồm các ấn phẩm từ thế kỷ 19 đến năm 1954 và một số tài liệu từ thế kỷ 16,17 và 18, ví dụ như *DelFhistoria della China (1586)*, *Dic- tionarivm annamiticvm lvsitanvm, et latinvm ope sacrae congregationis de propaganda fide in lvcem editvm AB / Alexandre de Rhodes (1651)*, *Voyage du Siam (1666)*, *Le grand dictionnaire historique, ou le mélange curieux de l'histoire sacrée et profane/ Louis Moréri (1759)*, *Voyage au Tonkin (1788)*; và các ấn phẩm thời kỳ 1954-1975. Trong vốn tài liệu đặc biệt này phải kể đến Kho tài liệu Đông dương với 20.461 nhan đề sách, 17.151 số báo, tạp chí và 78 bản đồ về đa dạng các chủ đề, thực sự là một kho tư liệu quý giá về một thời kỳ lịch sử của Việt Nam.

Quyết định 540.000 trang cho việc số hóa chỉ là một lựa chọn nhỏ (khoảng 5 %) của vốn tài liệu đặc biệt được đưa vào dự án. Với số lượng này thì rất dễ dàng để lựa chọn tài liệu không thuộc phạm vi tác động của Luật bản quyền tác giả, nhưng lại khó khăn cho các quyết định liên quan đến nội dung trong rất nhiều ứng viên sáng giá. Vì vậy cần phải có các tiêu chí lựa chọn.

Các tiêu chí lựa chọn này phải phù hợp với mục đích của Dự án, đồng thời phải xuất phát từ chính sách số hóa của Thư viện. Dựa vào đó, ban quản lý Dự án, cán bộ chuyên môn của Thư viện và nhất là các “độc giả nặng kí” đã thảo luận và thống nhất như sau:

- Các ấn phẩm bằng tiếng Pháp của Việt Nam trước năm 1954
- Có giá trị lịch sử, văn hóa, khoa học và thẩm mỹ
- Số lượng bản in hiện còn giữ được tại Thư viện
- Tình trạng bảo quản tài liệu hiện thời
- Tài liệu hư hỏng và có nguy cơ mất đi vĩnh viễn
- Khả năng đáp ứng nhu cầu sử dụng 125 cuốn sách; 11 nhan đề/520 cuốn/ 9.580 số tạp chí; 10 bản đồ đã được lựa chọn theo các tiêu chí trên đây.

Nội dung của các tài liệu này liên quan đến ba nước Việt Nam, Lào và Campuchia

Phần kỹ thuật số hóa được giao cho Dirox, một công ty phát triển phần mềm của Pháp tại Việt Nam thực hiện. Ban quản lý dự án VALEASE hoạt động với tư cách là điều phối viên giữa nhiều cơ quan tại ba nước Việt Nam, Lào, Campuchia của dự án lớn và giữa các cơ quan thụ hưởng và công ty cung ứng kỹ thuật trong từng dự án cụ thể. Thư viện đã chủ động tham gia như một đối tác thực hiện dự án hơn là một cơ quan hưởng thụ để đảm bảo chất lượng sản phẩm và các hoạt động của dự án cho phù hợp với chính sách bảo quản tài liệu của mình. Điều đó đòi hỏi Thư viện và các đối tác phải duy trì việc quản lý và kiểm soát đầy đủ trong suốt quá trình thực hiện. Được tiến hành song song với hoạt động thường xuyên của Thư viện, Dự án này không được gây ra bất kỳ thay đổi nào trong các mối quan hệ và các thủ tục trong cơ quan. Sự liên lạc giữa các đối tác tham gia dự án là một điểm quan trọng khác đã được đề cập ngay từ khi bắt đầu dự án.

1. QUẢN LÝ VÀ KIỂM SOÁT ĐẦY ĐỦ TRONG QUÁ TRÌNH THỰC HIỆN

2. Chuẩn bị tài liệu

Tại thư viện, Nhóm chuẩn tài liệu cho việc số hóa được thành lập gồm 4 người là Trưởng phòng và Phó phòng Tổ chức Kho tài và 2 cán bộ của Phòng Bảo quản/Tu bổ tài liệu. Việc đầu tiên của nhóm này là tham quan vị trí số hóa của Dirox để đảm bảo an toàn và an ninh cho tài liệu của Thư viện trong quá trình số hóa. Tiếp theo là thống nhất lịch trình

giao nhận tài liệu và các thủ tục bảo quản tài liệu. Tài liệu sẽ được chuẩn bị và giao theo đợt khoảng 50 cuốn để số hóa trong vòng một tháng. Tài liệu quý được giao với số lượng ít hơn, thậm chí là một tài liệu mỗi đợt. Công việc chuẩn bị tài liệu cơ bản gồm: 1) Chọn tài liệu theo tiêu chí đã thống nhất; 2) Kiểm tra tình trạng vật chất của tài liệu liên quan đến các yếu tố như số trang có mạch lạc và đầy đủ? In ấn có bị xiên xẹo? Lề có nhỏ hơn 1cm? Có phải thay thế định dạng tài liệu trước khi số hóa không (ví dụ như tháo rời, bó khung)..3) Thực hiện các sửa chữa nhỏ (cho mục đích số hóa đúng); 4) Làm danh sách, đóng gói và chuẩn bị vận chuyển tài liệu. Tất cả các công việc này rất quan trọng cho việc hỗ trợ quá trình số hóa liên quan đến thời gian và chất lượng. Ngoài ra, Nhóm chuẩn bị còn phải nắm được các thông tin khác của tài liệu như sự đa dạng của mật độ trên một trang, các trang bị mờ, tỉ lệ phần trăm minh họa cao... để chủ động về thời gian chờ tài liệu được số hóa.

Chuyên môn về tình trạng vật chất của vốn tài liệu được tích hợp vào qui trình thực hiện của dự án số hóa. Kiến thức và kinh nghiệm của Nhóm chuẩn bị tài liệu trong việc chăm sóc tình trạng vốn tài liệu và các thủ tục chuẩn bị, đóng gói và vận chuyển được xây dựng và thực hiện tại Thư viện nhiều năm đã được tái sử dụng cho Dự án này là một thuận lợi lớn.

2. SỐ HÓA

Công việc của Dirox bao gồm scan, xử lý OCR, xây dựng công cụ tìm kiếm.

Mặc dù các bên tham gia dự án đã hiểu rõ mục đích của Dự án và Dirox là một công ty có kiến thức và kinh nghiệm về lĩnh vực số hóa, nhưng một số yêu cầu cụ thể của Thư viện vẫn cần được làm rõ trước khi thực hiện. Đó là việc bảo quản tài liệu trong quá trình số hóa, chất lượng của sản phẩm dạng số và việc tìm kiếm thông tin phải được dễ dàng cho người sử dụng. Việc bảo quản dữ liệu số cũng được Thư viện lưu ý với Cty ngay từ khi bắt đầu số hóa.

Dự án số hóa này không chỉ nhằm bảo quản nội dung tài liệu Pháp ngữ cổ mà còn phải giữ lại tài liệu nguyên bản sau khi số hóa. Đối với nhiều nhà nghiên cứu thì bản thân tài liệu cũng là đối tượng nghiên cứu, họ quan tâm đến cả kỹ thuật in, loại giấy, bìa và kỹ thuật đóng. Một điểm cần lưu ý nữa là giấy được sản xuất vào giữa thế kỷ 19 có chất lượng rất kém. Vì vậy, Cty đã được yêu cầu giảm thiểu tối đa các ảnh hưởng bất lợi cho tài liệu trong qui trình scan. Máy scan i2S đã được sử dụng trong Dự án này, cho phép hạn chế các thao tác trên sách. Sách được mở 120o hoặc tối đa 180o để tránh hư hỏng gáy và phần đóng. Ánh sáng nhân tạo có hại cho tài liệu cũng không được sử dụng. Các thủ tục về xử lý và sử dụng đúng cách khác như lưu giữ trong phòng có điều hòa, đeo găng tay cotton... cũng được thực hiện đầy đủ.

Chất lượng sản phẩm số không phải là vấn đề của nhà cung ứng mà là vấn đề của người sử dụng. Chất lượng cao của các tập tin phân phối có đáp ứng cho các nhu cầu khác nhau của người sử dụng hay không, đó là điều mà Thư viện phải đặt vấn đề rất rõ ràng. Trong trường hợp này, Thư viện muốn ngoài việc “đọc” tại chỗ người sử dụng được quyền: in tài liệu ra giấy hay chép vào USB; CSDL này cũng phải truy cập được từ bên ngoài, tải xuống PC, tải xuống một mạng LAN hay WAN. Trong một số trường hợp Thư viện còn cần in lại tài liệu như bản gốc từ dữ liệu số hóa này. Dung hòa giữa chất lượng và dung lượng là vấn đề mà cả Thư viện và nhà cung ứng đều phải quan tâm. Trong khi còn nhiều điểm cần phải cân nhắc thì dự án vẫn phải tiến hành theo kế hoạch. Dirox thực hiện số hóa theo dạng ảnh và sử dụng phần mềm phục chế nhằm khắc phục những khiếm khuyết của tài liệu như vết ố, mờ. Các tài liệu số hóa dạng ảnh này được chuyển đổi sang dạng văn bản.

2. Truy cập, trình bày, bảo quản dạng số

Các tập tin — một khi đã qua chế độ đảm bảo chất lượng — thì được tiếp nhận vào môi trường trình bày và lưu trữ của thư viện. Phải tạo ra các chỉ mục cho tất cả các tập tin và phối hợp hàng ngàn các tập tin văn bản. Việc lập chỉ mục trong Dự án này được thực hiện đơn giản hơn yêu cầu của Thư viện nhưng cũng cho phép tìm kiếm theo tác giả, nhan đề và chú đề. Việc truy cập cũng mới chỉ hạn chế cho những người sử dụng đến thư viện. Người sử dụng có thể “đọc” tại chỗ, in ra giấy hoặc chép vào USB những phần cần thiết. Sản phẩm sẽ được đưa ra ở dạng CD như kế hoạch ban đầu. Bên cạnh đó, Thư viện đã đề nghị và được Ban quản lý dự án cấp thêm một máy chủ để lưu giữ cho cả mục đích sử dụng theo mong muốn nêu trên và cho việc bảo quản dữ liệu dạng số.

Một triển lãm rất đẹp vào ngày ra mắt sản phẩm số là cách rất tốt để quảng bá đến người sử dụng về một kho tàng mang giá trị khoa học và nghệ thuật dưới dạng số hóa đã có tại Thư viện Khoa học Tổng hợp Thành phố Hồ Chí Minh. Công trình “Kỹ thuật của người An Nam” của Henri Orger đã thu hút được sự quan tâm của những nhà nghiên cứu trong nước và nước ngoài. Nó không chỉ được lưu giữ ở dạng số mà còn được tiếp tục như một dự án mới: in lại tác phẩm này, trong đó phần lời chú thích bằng tiếng Hán và Nôm đã được dịch sang tiếng Việt, Anh và Pháp.

• KẾT LUẬN

Dự án đã thành công trong việc đạt được mục tiêu đề ra là có được một sản phẩm dữ liệu số. Dự án này đã cung cấp cho thư viện kinh nghiệm về nhiều mặt của việc số hóa đại trà và cách thức phối hợp với các cơ quan khác nhau trong quá trình thực hiện. Thư viện cũng rất ý thức về việc bảo quản dữ liệu số, giữ đúng chất “lưu trữ” chứ không đơn thuần là “lưu giữ”. Việc này phải được quan tâm ngay từ khi bắt đầu dự án bằng việc lựa chọn các đối tượng số hóa giá trị nhất để chuyển tải thông tin cho tương lai. Đảm bảo các đối

tượng số hóa được thực hiện ở đây có thể đọc được ở nơi khác, nhưng gì đang làm hôm nay có thể sử dụng được ngày mai. Nói chung các nỗ lực bảo quản dữ liệu số kể cả việc quản lý sự thay đổi công nghệ sao cho các đối tượng số hóa này được phù hợp là một nỗ lực lớn đòi hỏi việc lập kế hoạch tiếp tục của Thư viện, trong đó việc chuyển dự án lên chương trình, sáp nhập vào môi trường lớn hơn là những vấn đề đáng quan tâm.