

THƯ VIỆN VATICAN VÀ DỰ ÁN BẢO TỒN BỘ SƯU TẬP CỔ QUÝ HIẾM

Trần Thị Kiều Nga

Viện Thông tin Khoa học xã hội

1. Lịch sử và nguồn tư liệu cổ, quý, hiếm của Thư viện Vatican

1.1. Lịch sử Thư viện Vatican

Thư viện Tòa thánh Vatican, thường gọi là Thư viện Vatican, được thành lập năm 1475 và được xem là một trong số những thư viện cổ xưa nhất thế giới cho đến nay. Trước đó, năm 1451, giáo hoàng Nicholas V đã thành lập thư viện của riêng mình với số tài liệu ban đầu gồm 350 bản bằng tiếng Hy Lạp, tiếng Hebrew và tiếng Latin cổ. Đến năm 1455, thư viện đã có 1.200 tài liệu, trong đó có 400 cuốn bằng tiếng Hy Lạp cổ đại. Cũng năm đó, Giáo hoàng Nicholas V mất. Việc tiếp tục sưu tầm và tích lũy tư liệu bị gián đoạn. Giáo hoàng Sixtus đệ IV đã tiếp tục công việc của người tiền nhiệm. Năm 1475, ngài chính thức thành lập Thư viện Tòa thánh Vatican với tên ban đầu là Thư viện Palatine. Giáo hoàng Sixtus đệ IV đã cử Batolomeo Platina, một nhà văn và thiên văn học nổi tiếng thời kỳ Phục Hưng người Italia, làm quản thủ thư viện đầu tiên. Công việc đầu tiên Batolomeo tiến hành là lập danh mục toàn bộ tài liệu của thư viện. Trong năm 1475, ông đã hoàn thành bản danh mục của mình với 3.500 thủ bản, bao gồm tài liệu chép tay, các tài liệu thuộc nhiều lĩnh vực khác nhau. Vào thời điểm bấy giờ, bộ sưu tập này được xem là bộ sưu tập lớn nhất Châu Âu. Tiếp sau đó, Giáo hoàng Julius II đã cho mở rộng thêm toà nhà của Thư viện và tiếp tục phát triển thư viện. Năm 1587, Giáo hoàng Sixtus đệ V đã chỉ định kiến trúc sư Domenico Fontana xây một toà nhà mới cho thư viện, đặt ngay tại Thánh đường Vatican. Toà nhà ấy vẫn được sử dụng cho đến ngày nay và chính là Thư viện Vatican bây giờ. Trong suốt thời kỳ

Chấn hưng Công giáo, Thư viện Vatican đã giới hạn việc khai thác tài liệu và lập ra danh mục sách cấm, đặc biệt đối với những người theo đạo Tin Lành. Đến năm 1883, khi Giáo hoàng Leo XIII tại vị, lệnh cấm này mới chấm dứt và Thư viện được mở cửa phục vụ đại trà cho đến nay.

1.2. Bộ sưu tập của Thư viện Vatican

Bộ sưu tập của Thư viện Vatican được xem là một trong những bộ sưu tập cổ, quý hiếm nhất thế giới hiện nay. Bộ sưu tập có khoảng 75.000 bản chép tay về tôn giáo; 1,1 triệu sách in; 8.500 sách incunabula, 330.000 tài liệu khác gồm cả microfilm bằng tiếng La Mã, Hy Lạp, tiếng Hebrew, tiếng Ethiopia cùng các loại huân chương, huy chương và tiền đồng qua các đời giáo hoàng, các tác phẩm nghệ thuật. Phần lớn tài liệu trong số đó là các bản chép tay. Mỗi năm Thư viện được bổ sung khoảng 6.000 tài liệu thông qua việc được tặng thừa kế hoặc bổ sung theo diện mua, sưu tầm, biếu, tặng... Những tài liệu lâu đời nhất có niên đại từ thế kỷ thứ nhất. Tài liệu cổ nổi tiếng nhất trong thư viện chính là Bộ luật Codex Vaticanus Graecus 1209 của Kinh thánh. Năm 1623, Thư viện Giáo hoàng nhận được 3.500 bản chép tay như là quà cảm ơn của Công tước xứ Bavaria. Nội dung của bộ sưu tập bao gồm các bộ kinh thánh, các sưu tập những bài viết về luật lệ và phép tắc tôn giáo và các sách thế tục khác. Tổng số mét giá để xếp tài liệu của Thư viện Vatican có chiều dài 42km.

Hằng năm, Thư viện Vatican có 200 học giả tiếp cận tài liệu trực tiếp và có từ 4.000-5.000 học giả tới khai thác tài liệu thông qua các hoạt động hợp tác, nghiên cứu.

2. Dự án số hóa tư liệu quý hiếm ở Thư viện Vatican

Mục tiêu của số hóa các tài liệu quý hiếm ở Thư viện Vatican là hỗ trợ truy cập mở tới các di sản của nhân loại. Số lượng các tài liệu quý hiếm được dự kiến số hóa là 80.000 thủ bản (codex) từ thời kỳ Trung Cổ đến giai đoạn chủ nghĩa nhân văn trong thời kỳ Phục Hưng.

Tiêu chí để lựa chọn tài liệu ưu tiên số hóa như sau:

- Tinh xảo, dễ hỏng, có nguy cơ mất thông tin;
- Quan trọng và quý hiếm;
- Lựa chọn theo các dự án và tài trợ;
- Theo nhu cầu của các học giả.

Các nhà thư viện ở Thư viện Vatican đã mất rất nhiều năm để phân loại các thủ bản theo các tiêu chí trên. Bên cạnh đó, nhiều tài liệu trong bộ sưu tập được dát thêm vàng và bạc nên chúng sẽ bị phá hủy dưới ánh sáng của đèn quét. Vì thế, việc số hóa phải được thực hiện thành một quy trình chuyên biệt và có sự điều chỉnh ánh sáng một cách cụ thể đối với từng tài liệu. Thêm nữa, các thiết bị quét cũng phải được điều chỉnh theo đúng những đòi hỏi của các nhà lưu trữ ở Vatican, các tài liệu được đánh giá về tình trạng lưu giữ rất cẩn trọng trước và sau khi số hóa.

Ngoài những tiêu chí lựa chọn về tài liệu, những tiêu chí về công nghệ cũng quan trọng không kém. Việc thiết lập dự án số hóa dựa trên những tiêu chuẩn mở về siêu dữ liệu và các APIs (Application Programming Interface Set), một loại giao diện để các phần mềm và hệ điều hành kết nối và làm việc tương thích, cho phép các bộ sưu tập số hóa tương tác với nhau. Thêm vào đó, việc lựa chọn IIIF (International Image Interoperability Framework), thực chất gồm một bộ các APIs kết hợp, đưa vào sử

dụng nhằm mục đích sau:

- Giúp các học giả có thể truy cập tới các tài liệu dạng ảnh một cách đồng nhất và hết sức đa dạng;

- Xác định giao diện của các chương trình ứng dụng thông thường dùng để hỗ trợ tương tác giữa các bộ lưu trữ ảnh;

- Phát triển và đi sâu vào các công nghệ chia sẻ tài liệu như: kinh nghiệm của người dùng tin trên thế giới trong việc xem ảnh, thao tác ảnh, chú thích ảnh, so sánh ảnh.

Năm 2012, sau quá trình tìm hiểu, đánh giá và lựa chọn, xây dựng quy trình, dự án số hóa ở Thư viện Vatican chính thức bắt đầu. Dự án số hóa tư liệu quý hiếm được thực hiện trên cơ sở hợp tác giữa Thư viện Vatican và Thư viện Bodleian, một trong số những thư viện cổ nhất Châu Âu thuộc Đại học Oxford. Dự án này được Quỹ Polonsky, có trụ sở ở London, tài trợ. Dự án này tập trung vào ba nhóm tài liệu chính: bản thảo tiếng Hê-bơ-rơ, bản thảo Hy Lạp, và tài liệu in thời kỳ ban đầu (inculabula). Các nhóm tài liệu được chọn vì tầm quan trọng về học thuật của chúng và vì độ dài của bộ sưu tập trong cả hai thư viện. Dự án này kéo dài trong bốn năm với kết quả số hóa được 1,5 triệu trang tài liệu được lựa chọn.

Vào ngày 20/3/2014, Giáo hoàng đã long trọng tuyên bố Dự án số hóa tiếp nối dự án đang thực hiện đã được ký kết giữa Vatican với tập đoàn NTT Data Cooperation, Nhật Bản. Theo đó, tập đoàn NTT sẽ tài trợ cho Dự án các kỹ thuật viên và thiết bị trị giá khoảng 18 triệu Euro. Đáng chú ý, Dự án có thể số hóa lên tới 79.000 tài liệu và giai đoạn 1 của Dự án kéo dài trong 4 năm. Các tài liệu số hóa sẽ được lưu trữ trong ba máy chủ, mỗi máy chủ có dung lượng lên tới Petabyte (1 triệu tỷ bytes). Các máy chủ sẽ do tập đoàn Dell EMC cung cấp.

3. Một số ứng dụng công nghệ số hóa

Do quy mô và mục tiêu của các nhà quản thủ Thư viện Vatican nên các ứng dụng công nghệ cho Dự án số hóa được lựa chọn kỹ càng. Thư viện số của Vatican được chia thành sáu mục lục như sau: Mục lục tài liệu viết tay, Mục lục các tài liệu lưu trữ, Mục lục sách in nói chung, Mục lục tài liệu Incunabula, Mục lục tài liệu đồ họa, Mục lục về tiền đúc và huy chương. Các mục lục tương tác được với các biểu ghi theo định dạng MARC 21, EAD (Encoded Archival Description) và TEI-P5 (Text Encoding Initiative_Part 5), siêu dữ liệu dành riêng cho thủ bản chữ Ả Rập ở Vatican. Do đó, việc tìm kiếm một hệ thống quản trị và phổ biến cho các tài liệu ở Thư viện Vatican rất phức tạp, có sự liên kết của nhiều dạng siêu dữ liệu.

Hệ thống dịch vụ lưu trữ số dành cho thư viện và bảo tàng AMLAD của tập đoàn NTT, Nhật Bản được Thư viện Vatican lựa chọn là một hệ thống cho phép lưu trữ và phổ biến các sách in, tài liệu viết tay, tài liệu đa phương tiện có giá trị chuyên biệt cho các tổ chức như thư viện, phòng trưng bày, trung tâm lưu trữ, bảo tàng. AMLAD lưu trữ siêu dữ liệu theo kiểu sơ đồ hóa của riêng hệ thống, kết nối các siêu dữ liệu khác lại với nhau. Ví dụ như, AMLAD liên kết “Tên sách” từ tài khoản Library Books với “Chủ đề” từ tài khoản Archives. Đồng thời AMLAD cũng cho phép người dùng tin tìm kiếm chính xác các thông tin được lưu trữ dưới nhiều định dạng khác nhau nhưng có liên quan tới nhau thông qua cơ chế tìm kiếm chéo. Thậm chí AMLAD cũng cho phép người dùng tin tìm kiếm các tập tin ở nhiều định dạng khác nhau theo kiểu thủ công. Thêm vào đó, việc sử dụng công nghệ Watermark, công nghệ mã hóa kiểm soát bản quyền tài liệu số, giúp ngăn chặn người dùng tin thao tác sai mục đích như việc sao chép những ảnh đã được giới hạn sử

dụng. AMLAD cũng hỗ trợ cho việc sử dụng các giao diện của những thiết bị hiển thị khác nhau như điện thoại thông minh, máy tính bảng, máy tính cá nhân, xem phim hoặc ảnh có JPEG 2000. Hệ thống này cũng sử dụng các ứng dụng đọc tài liệu số giúp người dùng tin xem, đọc tài liệu như tài liệu thông thường. AMLAD cung cấp cho người dùng tin nhiều phương pháp tìm kiếm thông tin khác nhau như tìm theo từ khóa, tìm theo phân loại, tìm theo tài liệu cụ thể, tìm theo những định hướng xác định dưới dạng câu hỏi “What you mean?”. Cuối cùng, AMLAD cũng tương thích với các tiêu chuẩn quốc tế về lưu trữ tài liệu như: tiêu chuẩn Hệ thống thông tin lưu trữ mở (OAIS- Open Archival Information System), mô hình tham chiếu (ISO 14721:2012) đối với hệ thống thông tin lưu trữ mở, chuẩn siêu dữ liệu Dublin Core, chuẩn truyền dẫn và mã hóa siêu dữ liệu (METs-Metadata Encoding and Transmission Standard), giản đồ mô tả thư mục đối tượng siêu dữ liệu cho tiêu chuẩn truyền dẫn và mã hóa siêu dữ liệu (MODs- Metadata Objects Description Schema) dựa trên XML vốn được Thư viện Quốc hội Mỹ phát triển từ năm 2002, chuẩn mô tả tài liệu lưu trữ (EAD) dưới cú pháp XML dành cho việc lưu các trợ giúp tìm kiếm thông tin do Hội Lưu trữ Mỹ phát triển. Ngoài ra, AMLAD cũng hỗ trợ các giao thức kết nối chuẩn như: OAI-PMH (Opening Archive Initiative Protocol for Metadata Harvesting) và SRU/SRW (Search/Retrieve via URL- Search/Retrieve Web service). Nhờ những tiện lợi và khả năng tạo sự tương thích rộng như đã kể trên mà hệ thống AMLAD của tập đoàn NTT được lựa chọn sử dụng cho việc tìm kiếm, lưu trữ và phổ biến thông tin trong các cơ sở dữ liệu số của thư viện Vatican.

Thư viện số Vatican sử dụng định dạng FITS (Flexible Image Transport System) trong số hóa các ảnh và các tài liệu có ảnh. FITS là

một hệ thống định dạng dùng cho việc truyền dẫn các hình ảnh linh hoạt được Cơ quan hàng không vũ trụ Mỹ (NASA) phát triển cho các bức ảnh thiên văn của họ. Đây là lần đầu tiên, FITS được sử dụng cho số hóa và lưu trữ các thủ bản. Hiện nay, FITS đã được NASA phát triển tới phiên bản (version) 3.0. Một file FITS được tạo nên từ các biểu ghi dung lượng 2880 - bytes, còn được gọi là khối FITS. Khối FITS phân chia thành 2 phần là phần tiêu đề và phần dữ liệu. Đặc trưng chính của định dạng FITS là có thể lưu trữ các siêu dữ liệu ảnh theo tiêu đề chính mã ASCII (American Standard Code for Information Interchange_II/Bộ mã ký tự). Điều này giúp cho người dùng tin có thể kiểm tra tiêu đề theo nguồn. Mỗi file FITS bao gồm một hoặc hơn một tiêu đề có gắn các thẻ của ảnh theo dạng mã ASCII 2 chứa đựng các cặp từ khóa/giá trị nằm xen kẽ giữa các dữ liệu. Các cặp từ khóa/giá trị cung cấp thông tin về kích cỡ, nguồn gốc, tọa độ, định dạng hệ nhĩ phân, lịch sử dữ liệu... trong khi đó, nhiều từ khóa cùng lúc cũng có thể dùng cho một FITS. So với các định dạng file khác, ví dụ như so với định dạng TIFF, FITS không bị giới hạn cho việc đọc các tập tin theo mọi kích thước và mọi loại số hoặc dữ liệu văn bản có thể được lưu: Số nguyên hoặc số thực, 32 hoặc 64 bit, hình ảnh được xử lý như các ma trận hai chiều trong đó màu sắc và mã hóa ảnh được xem là các giá trị của ma trận. Mặt khác, các chương trình xử lý ảnh như GIMP, Photoshop, XnView và IrfanView có thể đọc các hình ảnh FITS đơn giản, nhưng thường không thể giải thích các bảng phức tạp hơn. Trong dự án số hóa ở Thư viện Vatican, việc chọn tệp FITS ở định dạng để bảo quản lâu dài cho thấy FITS được lựa chọn không chỉ dành cho công chúng đọc mà còn dành cho lưu trữ dữ liệu.

Vì hai CSDL trực tuyến của thư viện Vatican là CSDL Thủ bản và CSDL Tài liệu lưu trữ có

mối liên kết phức tạp giữa các siêu dữ liệu khác nhau nên chúng được quản lý theo hai khổ mẫu dữ liệu là TEI-P5 và EAD theo cú pháp XML. Tuy là hai sưu tập dữ liệu chuyên biệt nhưng chúng lại cùng sử dụng một ứng dụng là InforMA. InforMA là tên của một ứng dụng được thiết lập bởi một công ty phần mềm AG của Đức, cung cấp hỗ trợ toàn diện cho CSDL theo XML, lưu trữ nội dung theo XML, cung cấp các khả năng xử lý hồi cố dựa trên tiêu chuẩn XQuery (XML Query) và Xpath (XML Path), tiêu chuẩn về ngôn ngữ cho các nội dung đa phương tiện có cấu trúc hoặc không có cấu trúc, tương thích với Microsoft Office hoặc những sản phẩm dựa trên WebDAV (Web-based Distributed Authoring and Versioning- Một dạng giao thức trong môi trường World Wide Web). Những đặc trưng của InforMA được phần mềm mang tên Tamino hỗ trợ, một phần mềm do công ty phần mềm AG của Đức cung cấp.

Một ứng dụng khác được sử dụng trong Dự án là V-Smart/Iguana. Ứng dụng này, bao gồm hai ứng dụng kết hợp với nhau có tên V-smart và Iguana, giúp người dùng tin tham vấn các khả năng tìm kiếm thông tin và sử dụng thông tin trong CSDL tích hợp chung. Cấu trúc công nghệ của ứng dụng này dựa trên giao thức OAI-PMH và các yếu tố ký tự dành cho truy xuất dữ liệu (từ InforMA) tới nhập dữ liệu XML (tới V-Smart). Từ mỗi biểu ghi thư mục, V-Link (bộ giải URL mở) có thể tìm kiếm hoặc truy cập tới các nguồn tin và hồi cố những kết quả tìm có liên quan. Ứng dụng này cũng hỗ trợ cho các URIs (Uniform Resource Identifier Set/Chuỗi ký tự dùng để định danh tên, tài nguyên trên Internet) có liên quan tới giao diện trình bày Web của các tài liệu inculabula được số hóa.

Trong khi hai CSDL thư mục trực tuyến trong số 06 thư mục CSDL trực tuyến của dự án số hóa thư viện Vatican sử dụng dạng mã

hóa các chuyên biệt (CSDL thư mục thủ bản sử dụng TEI-P5 và CSDL thư mục tài liệu lưu trữ sử dụng EAD) thì bốn CSDL thư mục trực tuyến còn lại sử dụng khổ mẫu MARC 21. Phần mô tả các thủ bản xác định một yếu tố chuyên biệt để cung cấp thông tin mô tả cơ bản về tài liệu chép tay. TEI-P5 được phát triển nhằm thỏa mãn các yêu cầu của các nhà thư mục học và học giả liên quan tới các thủ bản Châu Âu theo phong cách truyền thống. Hiện nay, TEI-P5 đã được mở rộng dành cho các loại tài liệu khác cũng có thể bao gồm các tài liệu khắc thạch bản. TEI-P5 cung cấp khung thống nhất về xác định ngôn ngữ đánh dấu, cách đơn giản để tổ chức các nguồn tin văn bản có cấu trúc. Một biểu ghi của TEI-P5 bao gồm các thông tin sau: Yếu tố mô tả thủ bản, Yếu tố cấp độ nhóm, Định danh tài liệu, Đề mục của thủ bản, nội dung thông tin của thủ bản, yếu tố mô tả vật lý, Lịch sử tài liệu, Thông tin thêm, Các phần của thủ bản, Mảnh của thủ bản, Phân đoạn mô tả tài liệu. Các thông tin trên được viết theo cú pháp của XML.

EAD là một chuẩn khổ mẫu mô tả các tài liệu lưu trữ trong đó các yếu tố thông tin được mã hóa theo cú pháp XML. EAD do Hội Lưu trữ Mỹ đưa ra. Phiên bản version 1.0 được bắt đầu từ năm 1998. Hiện nay EAD đã phát triển thành phiên bản EAD3. EAD được xem là một trong số những tiêu chuẩn về siêu dữ liệu cho các tài liệu lưu trữ hiện nay. Một biểu ghi theo chuẩn EAD3 bao gồm 165 yếu tố thông tin được đưa vào so với phiên bản ban đầu chỉ có 146 yếu tố thông tin. Các yếu tố trên dùng để mô tả chung toàn bộ tài liệu cũng như mã hóa các bảng kê theo nhiều cấp độ khác nhau. Các yếu tố có thể được sắp xếp theo bộ tiêu chuẩn nội dung như DACs (Tiêu chuẩn nội dung do Hiệp hội Lưu trữ Mỹ đưa ra), ISAD(G) (Tiêu chuẩn mô tả lưu trữ quốc tế do Hiệp hội lưu trữ quốc tế đưa ra) hoặc theo các tiêu chuẩn

về cấu trúc thông tin khác như MARC, Dublin Core nhằm làm tăng tính linh hoạt cũng như sự tương tác của dữ liệu. EAD phiên bản đầu tiên được viết dưới dạng cú pháp DTD, phiên bản EAD3 hiện nay được viết theo dạng cú pháp XML DTD.

Tóm lại, với những ứng dụng về công nghệ thông tin hiện đại nhất, quy mô dự án vô lớn cùng với sự quý hiếm và có giá trị khoa học cao của các thủ bản, dự án thư viện Vatican là một dự án về số hóa tài liệu được mong chờ nhất thế giới hiện nay. Với quy mô của dự án như thế này, các nhà quản thủ thư viện Vatican đang trên con đường bảo tồn và phát huy các thủ bản quý hiếm của mình có hiệu quả tối ưu.

TÀI LIỆU THAM KHẢO

1. Bridging between Cultural Heritage Institutions: 9th Italian Research Conference, IRCDL 2013, Rome/Italia, January 31 - February 1, 2013: *Revise Selected Papers*, Edition Springers, 2013, 200 p.
2. Encoding Archival Description Tag Library
<https://www2.archivists.org/sites/all/files/TagLibrary-VersionEAD3.pdf>. Truy cập ngày 14/06/2017.
3. It FITS the Cultural Heritage
http://www.unesco.org/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Michetti_Manoni_28_C_1110.pdf. Truy cập ngày 14/06/2017
4. TEI-P5 : Guidelines for electronic Text Encoding and Interchange
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>. Truy cập ngày 15/06/2017.
5. The tech behind digitizing the Vatican Library : Open access for all | #EMCWorld
<https://siliconangle.com/blog/2014/05/23/the-tech-behind-digitizing-the-vatican-library-open-access-for-all-emcworld/>. Truy cập ngày 14/06/2017