

DSPACE – Giải pháp tạo lập, lưu trữ và phổ biến tài nguyên điện tử cho các thư viện ở Việt Nam

Nguyễn Huy Chương¹⁰, Nguyễn Tiến Hùng¹¹

Tóm tắt: Bài viết giới thiệu tổng quan về phần mềm thư viện số Dspace, mô tả các chức năng và thiết kế của hệ thống thư viện số mã nguồn mở Dspace, và cách tiếp cận các vấn đề khác nhau trong thư viện số và thiết kế lưu trữ.

I. GIỚI THIỆU

Tháng 3 năm 2000, Công ty Hewlett-Packard (HP) đã trao tặng 1.800.000 USD cho Thư viện MIT (Massachusetts Institute of Technology) trong khuôn khổ một chương trình hợp tác 18 tháng để xây dựng thư viện số Dspace, một kho lưu trữ năng động các định dạng kỹ thuật số các tài nguyên tri thức của các tổ chức nghiên cứu đa ngành. Một tháng sau khi giới thiệu, ngày 4/11/2000, HP Labs và Thư viện MIT đã phát hành trên toàn thế giới hệ thống Dspace theo các điều khoản của giấy phép mã nguồn mở BSD[1] như là một dịch vụ mới của thư viện MIT. Là một hệ thống mã nguồn mở, Dspace miễn phí cho các tổ chức khác để sử dụng, sửa đổi và mở rộng theo các yêu cầu của họ để đáp ứng nhu cầu của từng tổ chức.

Dspace ra đời là một nỗ lực để giải quyết một số vấn đề mà giảng viên của MIT đã gặp phải trong những năm qua. Khi giảng viên và các nhà nghiên cứu khác phát triển các tài liệu nghiên cứu và các ấn phẩm học thuật trong các định dạng kỹ thuật số ngày càng phức tạp, có một nhu cầu để thu thập, bảo quản và phân phối chúng: một công việc tốn thời gian và tốn kém cho các giảng viên cá nhân và các phòng ban của họ, phòng thí nghiệm, và các trung tâm. Dspace cung cấp một cách để quản lý các tài liệu nghiên cứu và các ấn phẩm trong một kho lưu trữ chuyên nghiệp để duy trì, cung cấp cho họ khả năng hiển thị lớn hơn và khả năng tiếp cận theo thời gian.

Đầu tiên Dspace được xây dựng theo hướng tiếp cận bề rộng: nó hỗ trợ tất cả các chức năng mà một tổ chức nghiên cứu cần có: một dịch vụ tạo lập tài nguyên số, kho

¹⁰ Tiến sĩ, Giám đốc Trung tâm Thông tin – Thư viện, Đại học Quốc gia Hà Nội

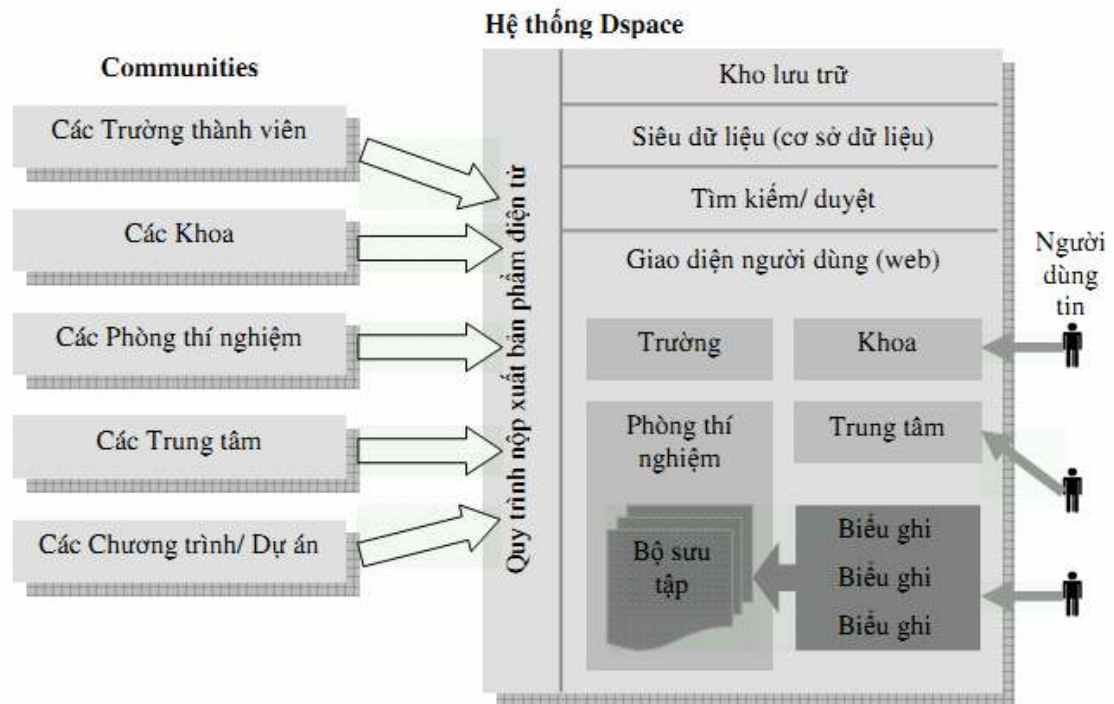
¹¹ Kỹ sư tin học, Giám đốc Công ty Phần mềm & Truyền thông VIC

lưu trữ tài nguyên số bằng một cách đơn giản nhất có thể với mục tiêu sẽ được ngay lập tức hữu ích tại MIT, dần dần có thể được mở rộng và cải thiện theo thời gian, và có thể phục vụ như một nền tảng cho các nghiên cứu trong tương lai.

II. TỔNG QUAN VỀ HỆ THỐNG DSPACE

1. Mô hình thông tin của hệ thống

Dspace được thiết kế để tạo thuận lợi cho các cá nhân có thể đóng góp các tài nguyên số vào hệ thống một cách dễ dàng. Mô hình thông tin của hệ thống được xây dựng xung quanh ý tưởng “Communities” tổ chức các đơn vị trực thuộc của một tổ chức nghiên cứu, một trường đại học có nhu cầu quản lý thông tin đặc biệt. Trong trường hợp của MIT (một trường đại học nghiên cứu lớn), “Communities” được định nghĩa là các trường thành viên, các khoa, phòng thí nghiệm, và các trung tâm của MIT. Mỗi “communities” có thể thích ứng với hệ thống để đáp ứng các nhu cầu cụ thể của đơn vị và quản lý quá trình nộp các xuất bản phẩm điện tử.



Hình 1: Mô hình tổ chức thông tin trong Dspace

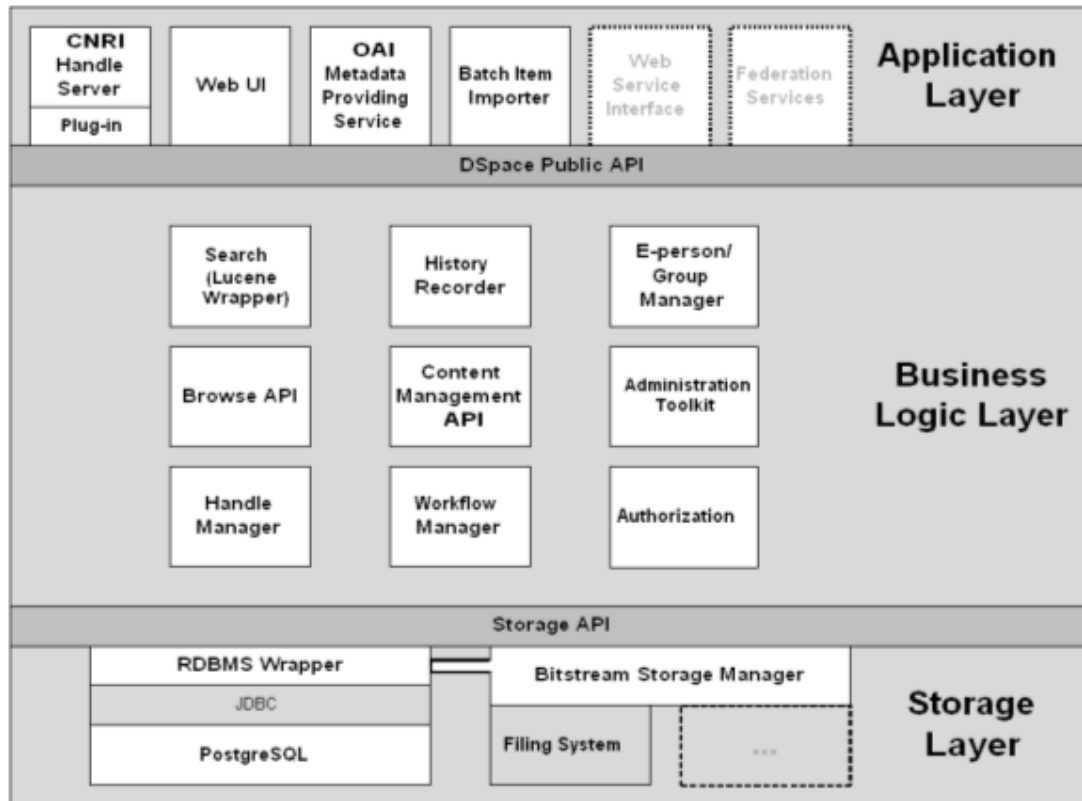
2. Kiến trúc hệ thống

Kiến trúc hệ thống của Dspace là một kiến trúc ba lớp:

- Lớp lưu trữ

- Lớp nghiệp vụ
- Lớp ứng dụng

Các lớp lưu trữ được thực hiện bằng cách sử dụng hệ thống tập tin, quản lý bởi các bảng trong cơ sở dữ liệu PostgreSQL. Lớp nghiệp vụ là lớp các chức năng cụ thể của Dspace, bao gồm cả các module luồng công việc, quản lý nội dung, quản trị, tìm kiếm và duyệt tài liệu. Mỗi module có một API để cho phép Dspace tùy chỉnh, nâng cấp các chức năng phù hợp với từng đối tượng. Cuối cùng, lớp ứng dụng bao gồm các giao diện cho hệ thống giao diện người dùng web và bộ xử lý theo lô, đặc biệt còn hỗ trợ OAI (*Open Archives Initiative: Sáng kiến lưu trữ mở*) và xử lý máy chủ để giải quyết định danh liên tục (Handle) đến các biểu ghi trong Dspace.



Hình 2: Kiến trúc hệ thống Dspace

3. Metadata (siêu dữ liệu)

Dspace sử dụng siêu dữ liệu chuẩn Dublin core để mô tả các thông tin về tài nguyên điện tử cần lưu trữ, phân phối. Trong đó có 3 yếu tố (thông tin) bắt buộc phải mô tả: Nhan đề, ngôn ngữ, ngày đăng, tất các yếu tố còn lại là tùy chọn. Ngoài ra, có một số các yếu tố bổ sung cho tài liệu: tóm tắt, từ khóa, siêu dữ liệu kỹ thuật và siêu dữ liệu quyền. Các siêu dữ liệu này được hiển thị trong biểu ghi của tài liệu trong hệ

thống Dspace và được lập chỉ mục để hỗ trợ tìm kiếm, duyệt thông tin trong hệ thống (duyệt theo Bộ sưu tập, theo chủ đề, theo các đơn vị thành viên của tổ chức...). Hệ thống hỗ trợ kết xuất siêu dữ liệu và tài liệu điện tử trong kho lưu trữ theo dạng chuẩn XML, và hiện đang phát triển để hỗ trợ chuẩn METS đối với các siêu dữ liệu kỹ thuật và siêu dữ liệu quyền cho các định dạng kỹ thuật số tùy ý.

4. Giao diện người dùng

Dspace sử dụng giao diện người dùng dạng web-based. Có 3 giao diện người dùng trong hệ thống Dspace: giao diện người dùng cho những người tham gia trong quá trình đăng xuất bản phẩm điện tử; giao diện cho người dùng tin: tìm kiếm, duyệt thông tin trong kho lưu trữ; giao diện người dùng cho người quản trị

Giao diện cho người dùng tin hỗ trợ tìm kiếm và nhận thông tin trả về bằng cách duyệt hoặc tìm kiếm siêu dữ liệu. Một biểu ghi thỏa mãn các điều kiện tìm kiếm trong kho lưu trữ sẽ được trả về, người dùng tin có thể tải tài liệu điện tử gắn với biểu ghi thông qua các siêu liên kết, người dùng tin có thể xem trực tiếp nội dung tài liệu trên web bằng cách cài đặt các plug-in cho trình duyệt web hoặc tải về máy tính của mình để xem thông qua các ứng dụng hỗ trợ đọc các định dạng điện tử khác nhau như: Microsoft Office, Acrobat, Windows Media Player, CAD/CAM ...

5. Luồng công việc trong Dspace

Bằng cách áp dụng mô hình luồng công việc, Dspace là một kho lưu trữ tài liệu điện tử mà nguồn mở đầu tiên đã giải quyết những vấn đề phức tạp của một thư viện khoa học tổng hợp tỉnh. Nói cách khác, mỗi thư viện quận, huyện trong tỉnh sẽ có những quy định rất khác nhau về các tài nguyên điện tử phải nộp cho “thư viện số”: loại tài liệu điện tử phải nộp là gì? Ai là người gửi tài liệu? Ai là người duyệt? Ai được xem và ai là người bị hạn chế xem các tài liệu này... Tất cả những vấn đề này đều được giải quyết bởi các đại diện của các thư viện quận, huyện và các cán bộ thư viện tỉnh. Sau đó được mô phỏng bằng luồng công việc cho mỗi bộ sưu tập để thực thi các quyết định. Mỗi thành viên trong Dspace đều được gán các quyền thích hợp với vai trò của mình: vai trò người đăng tải tài liệu điện tử, vai trò người biên tập siêu dữ liệu, vai trò quản trị các bộ sưu tập, vai trò quản trị hệ thống ...

Có 2 cách để tạo lập, quản trị bộ sưu tập số. Cách thứ nhất, thư viện có thể quy định tất cả mọi bạn đọc đều có quyền đăng tải tài liệu điện tử, và bất kỳ người dùng nào (trong nội bộ và bên ngoài) đều có quyền xem các tài liệu đã được đăng tải. Cách thứ

hai, thư viện có thể tổ chức mô hình lưu trữ và phân phối tài liệu chặt chẽ hơn: các tác giả nộp các tài liệu điện tử do mình tạo lập, sau đó sẽ có một bộ phận chịu trách nhiệm biên tập siêu dữ liệu và người có quyền cao nhất sẽ quyết định có xuất bản tài liệu đó không. Như vậy mỗi bước trong quá trình đăng tải tài liệu điện tử sẽ được xem xét, phê duyệt trước khi tài liệu đó được đưa vào bộ sưu tập, các tài liệu điện tử không được thông qua trong quá trình này sẽ không được phép lưu trữ trong hệ thống Dspace.

6. Nền tảng công nghệ

Dspace được thiết kế để chạy trên nền tảng UNIX, hiện nay đã có phiên bản chạy trên nền tảng hệ điều hành Windows. Các mã ban đầu được lập trình bằng ngôn ngữ Java. Các thành phần khác: Hệ quản trị CSDL sử dụng PostgreSQL; máy chủ web và Java Servlet sử dụng Apache và Tomcat; Jena - một bộ công cụ RDF được phát triển từ HP Labs, OAICat từ OCLC

III. MỘT SỐ TÍNH NĂNG NỔI BẬT CỦA DSPACE SO VỚI GREENSTONE

Hiện ở Việt Nam có 2 phần mềm thư viện số được sử dụng khá phổ biến là Dspace và GreenStone, dưới đây bài viết chỉ đề cập những tính năng nổi bật của Dspace với GreenStone để bạn đọc so sánh và có những lựa chọn phù hợp cho thư viện của mình.

Khả năng tùy chỉnh giao diện cao: Giao diện thống nhất chung cho tất cả các bộ sưu tập. Tất cả các thao tác đều thông qua web: Biên mục (đây đồng thời cũng là nhược điểm của Dspace vì việc biên mục trên web sẽ phụ thuộc vào sự ổn định của đường truyền, băng thông, thời gian xử lý ...), truy cập thông tin... Khi cần bổ sung tài liệu vào các bộ sưu tập không cần phải xây dựng lại từ đầu như Greenstone

Cấu trúc Bộ sưu tập trong Dspace khoa học hơn Greenstone. Dspace có cấu trúc các Bộ sưu tập theo nhiều cấp

Sử dụng hệ quản trị CSDL (PostgreSQL) độc lập nên đáp ứng tốt với Thư viện có số lượng tài liệu lớn

Khả năng phân quyền mạnh. Có thể phân quyền đến từng tài khoản người dùng, đến từng Bộ sưu tập hoặc thậm chí đến từng tài liệu. Các quyền được cấu hình khá chi tiết như: Quyền xem biểu ghi thư mục, Quyền xem toàn văn... Phần mềm Greenstone không làm được điều này

Hỗ trợ nhiều kiểu báo cáo: Lướt truy cập, lướt xem biểu ghi thư mục, lướt download...

Được Google Scholar hỗ trợ chỉ mục tài liệu

IV. MỘT SỐ HÌNH ẢNH TRIỂN KHAI DSPACE TẠI TRUNG TÂM TT-TV, ĐHQGHN



Hình 3: Minh họa màn hình tổ chức Đơn vị thành viên và Bộ sưu tập



Hình 4: Minh họa màn hình duyệt tài liệu theo Nhan đề



Hình 5: Minh họa màn hình duyệt tài liệu theo Tác giả



Hình 6: Minh họa màn hình duyệt tài liệu theo thời gian

V. KẾT LUẬN

Xây dựng thư viện số là xây dựng phương thức mới, công nghệ mới trong việc xử lý thông tin - tri thức. Đó là bảo quản, sưu tầm, tổ chức, quảng bá, và truy cập thông tin hay nói chính xác hơn là tri thức, tức là thông tin có ý nghĩa và hữu ích. Do đó, một thư viện số được xem như là nơi quản trị và cung cấp những bộ sưu tập thông tin có tổ chức. Với nguồn kinh phí hạn hẹp dành cho hoạt động thư viện hiện nay, chúng tôi cho rằng phần mềm nguồn mở Dspace là một giải pháp tốt để giải quyết những khó khăn và thách thức cho công tác tạo lập và quản lý các tài nguyên điện tử nội sinh không chỉ trong các thư viện đại học mà tại tất cả các hệ thống thư viện Việt Nam.

Danh mục tài liệu tham khảo

- [1] Berkeley Standard Distribution License, <<http://www.opensource.org/licenses/bsd-license.php>>.
- [2] DSpace, <<http://dspace.org>>.
- [3] Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>.
- [4] OAICat, <<http://www.oclc.org/research/software/oai/cat.shtm>>.
- [5] Handle System®, <<http://www.handle.net>>.