

THƯ VIỆN SỐ vỚI HỆ THỐNG NGUỒN MỞ

NGUYỄN MINH HIỆP, BA., MS.

GD. Thư viện Đại học Khoa học Tự nhiên
TP. Hồ Chí Minh

Hiện nay trên thế giới, xu thế phát triển Thư viện số đã trở thành một phần chủ đạo trong toàn cảnh hoạt động thông tin thư viện, trong khi phần mềm nguồn mở trở thành một hiện tượng toàn cầu. Giống như nhiên liệu và động cơ trong kỹ thuật, *nguồn mở* và *thư viện số* là hai yếu tố không thể tách rời.

Thư viện số

Từ năm 1994, một dự án xây dựng Thư viện số của Liên bang Hoa kỳ do bốn trường đại học lớn thực hiện đã khai sinh ra Thư viện số. Đã có nhiều định nghĩa khác nhau về Thư viện số, tuy nhiên đều thống nhất một mục tiêu chung xây dựng Thư viện số là nhằm đối phó với tình trạng ‘*quá tải thông tin – information overload*’ đồng thời tiến hành ‘*quản lý tri thức – knowledge management*’.

Tri thức là thông tin có ý nghĩa và hữu ích. Xây dựng Thư viện số là chọn lọc, sưu tầm, và tổ chức thông tin có ý nghĩa và hữu ích hay tri thức nhằm đáp ứng đúng yêu cầu của người sử dụng. Theo Branscomb, một nhà kinh tế tri thức thì ‘*Nếu ví thông tin như bột mì thì tri thức chính là bánh mì*’. Dựa vào lời ví von đầy ý nghĩa đó, ta có thể hiểu một cách nôm na rằng ‘*Xây dựng thư viện số là làm những chiếc bánh mì*’. Hay nói một cách khác, nếu chúng ta xây dựng thư viện số có nghĩa là chúng ta không phải chỉ phục vụ ‘*bột mì*’ tức là tài nguyên thông tin mình có, mà phải nhào nặn bột mì đó thành những

‘*chiếc bánh*’ theo yêu cầu và nhu cầu của người sử dụng. Nói một cách chuyên nghiệp, chúng ta phải xây dựng những **Bộ sưu tập kỹ thuật số** nhằm hình thành Thư viện số. Những định nghĩa có giá trị sau đây cho chúng ta thấy rõ ý nghĩa này.

1. Định nghĩa Thư viện số của Liên đoàn Thư viện số thế giới – DLF : *“Thư viện số là một tổ chức cung cấp tài nguyên, bao gồm cả đội ngũ chuyên nghiệp, để chọn lọc, cấu trúc, cung ứng truy cập, biên dịch, phân phối, bảo quản nguyên vẹn, và đảm bảo bền vững theo thời gian những bộ sưu tập kỹ thuật số để sẵn sàng phục vụ cộng đồng một cách kinh tế”*.

DLF (Digital Library Federation) là một liên hiệp thư viện và các cơ sở liên hệ có nhiệm vụ tổ chức việc sử dụng công nghệ thông tin điện tử nhằm mở rộng việc sử dụng các bộ sưu tập và phát triển dịch vụ.

2. Định nghĩa Thư viện số của TS. Ian Witten :

“Thư viện số là tập hợp những bộ sưu tập thông tin của các đối tượng số hoặc đã được số hóa có tổ chức và tập trung. Tập trung theo đề tài hay chủ đề và có tổ chức để thông tin dễ truy cập và lưu trữ theo những tiêu chuẩn chuyên biệt cung cấp hai khả năng chính:

- *P hương thức truy cập, chọn lọc, hiển thị tài nguyên số (dành cho người sử dụng);*

- **Phương thức xây dựng, tổ chức và lưu hành (dành cho cán bộ thư viện).**

TS. Ian Witten là chuyên gia Thư viện số của Đại học Waikato, New Zealand, chủ biên Phần mềm nguồn mở Thư viện số Greenstone.

Bộ sưu tập

Một bộ sưu tập thông tin bao gồm nhiều tài liệu dưới nhiều dạng thức khác nhau: văn bản, hình ảnh, âm thanh, hình ảnh động. Tài liệu là đơn vị căn bản từ đó sưu tập thông tin được xây dựng. Một bộ sưu tập có thể chứa nhiều loại tài liệu khác nhau. Ví dụ một bộ sưu tập về đề tài "Phổ cổ Hội An" sẽ bao gồm những tài liệu dạng văn bản về lịch sử, văn hoá, phong tục, vv...; tài liệu dạng hình ảnh về những di tích, trang phục, các bản thiết kế, vv...; tài liệu dạng âm thanh về những bài hát, dân ca, vv...; tài liệu dạng phim về những lễ hội, sinh hoạt cộng đồng, vv... Một bộ sưu tập chứa nhiều tài liệu với dạng thức khác nhau, tuy nhiên cung cấp một giao diện đồng nhất qua đó tất cả các tài liệu có thể được truy cập, mặc dù cách mà tài liệu đó hiển thị sẽ tùy thuộc vào phương tiện và dạng thức của tài liệu đó.

Một bộ sưu tập như thế trước khi trình bày phải qua một quá trình hình thành để tạo nên những cấu trúc hỗ trợ cho việc truy tìm và lướt tìm được dùng cho việc truy cập sưu tập. Khi xây dựng xong, bộ sưu tập có thể được xuất bản trên Internet hoặc xuất ra CD-ROM một cách hoàn toàn tự động. Một khi sưu tầm thêm tài liệu mới, ta có thể dễ dàng bổ sung thêm vào bộ sưu tập bằng cách tái xây dựng.

Một thư viện nói chung bao gồm nhiều bộ sưu tập khác nhau, mỗi sưu tập tổ chức mỗi khác, tuy nhiên hoàn toàn giống nhau về phương cách hiển thị. Những bộ sưu tập như thế có thể được tạo nên bằng một Phần mềm nguồn mở, chẳng hạn như Phần mềm

nguồn mở đa ngôn ngữ thư viện số Greenstone (Greenstone digital library multilingual open source software).

Hệ thống nguồn mở

Phần mềm nguồn mở cũng có một mối quan hệ cộng đồng. Mặc dù những định nghĩa về Phần mềm nguồn mở luôn chú trọng đến việc có thể sử dụng mã nguồn, điều chúng ta đáng lưu ý hơn là phần mềm nguồn mở được tạo nên và nuôi dưỡng bởi một cộng đồng. OSI (Open Source Initiative), đã nắm giữ thương hiệu *phần mềm nguồn mở*, là một tổ chức phi lợi nhuận nhằm thiết kế nên việc quản lý và xúc tiến việc xác định quyền hạn của nguồn mở đối với sản phẩm công cộng, đã có một định nghĩa về phần mềm nguồn mở như sau:

“Phần mềm nguồn mở không chỉ là một thuật ngữ cho biết việc có thể sử dụng mã nguồn, mà còn sử dụng những tiêu chí bao gồm việc tự do phân phối phần mềm ứng dụng, cũng như dự phòng cho việc sử dụng tương lai của những mã nguồn nhằm chỉnh sửa và thực hiện việc mở rộng ứng dụng mà không phân biệt đối xử giữa những người hay nhóm người sử dụng chúng”.

Phần mềm nguồn mở thường mang đặc tính như phần mềm tự do. Một sản phẩm phần mềm nguồn mở sẽ thông thường được phân phối với một giấy phép không chế việc sử dụng và phân phối sản phẩm đó. Richard Stallman, người sáng lập Dự án GNU (GNU là viết tắt của Gnu's Not UNIX) và là người đứng đầu trong thế giới phần mềm nguồn mở, cũng là người được giới Tin học Việt Nam đánh giá là “*người cộng sản của xã hội phần mềm*” nhân dịp ông viếng thăm Hà Nội vào tháng 3/2004, đã ấn định bốn loại tự

do cho những phần mềm ứng dụng nguồn mở được hỗ trợ cho việc cấp giấy phép là:

- ***Tự do chạy chương trình với bất cứ mục đích nào;***
- ***Tự do chỉnh sửa cho phù hợp với yêu cầu của mình;***
- ***Tự do tái phân phối bản sao để giúp người khác sử dụng;***
- ***Tự do phát triển chương trình và bán rộng rãi phần phát triển đó nhằm mang đến lợi ích chung cho cộng đồng.***

Những loại giấy phép thông dụng bao gồm:

- GNU General Public License (GPL)
- Creative Commons
- GNU Lesser Public License (LGPL), Artistic
- Berkley System Distribution (BSD) License, Apache Software License, MIT License, NCSA License
- Mozilla Public License (MPL hay MozPL), Netscape Public License (NPL)
- OCLC Research Public License

Giấy phép phần mềm nguồn mở được xem như là mối tương quan đặc thù giữa một hệ thống sắp được tái đóng gói để phân phối với những hệ thống khác sử dụng sản phẩm đó trong môi trường thương mại.

Vốn tài liệu thư viện, đội ngũ nhân viên thư viện, và ngay cả hạ tầng cơ sở, trang thiết bị thư viện thường được dùng để phục vụ rộng rãi cho cộng đồng người sử dụng dựa trên cơ sở phi lợi nhuận. Thư viện cũng giống như những tổ chức khác, người sử dụng thường xuyên phần mềm nguồn mở thông qua đội ngũ nhân viên thư viện thường không quan tâm đến bao nhiêu dịch vụ thư viện được cung cấp đã dùng những phần mềm nguồn mở nào.

Tham gia ứng dụng phần mềm nguồn mở nhiều nhất đối với hầu hết những người sử dụng máy tính là thông qua Internet, không chỉ bởi vì hạ tầng cơ sở này được xây dựng dựa trên những ứng dụng phần mềm nguồn mở, mà còn bởi vì Internet đã trở nên có mặt mọi lúc mọi nơi, điều này khiến cho những công nghệ nổi trội được dễ dàng công nhận và sử dụng rộng rãi. Hệ điều hành nguồn mở Linux là một ví dụ điển hình. Linux còn được xem như là một bích chương quảng bá phần mềm nguồn mở.

Thuật ngữ then chốt kết nối thư viện số với phần mềm nguồn mở là ***mở***. Thật vậy, thư viện số đôi khi được xem như là thư viện số mở (ODL – Open Digital Library), và mô hình “mở” chẳng hạn như lưu trữ mở đã nảy sinh ra trong chừng mực nào đó vấn đề chia sẻ quyền sở hữu trí tuệ. Nhiều người cho rằng, thư viện đã bắt nguồn từ cơ sở lý luận rằng một xã hội mở là phụ thuộc vào việc trao đổi tự do những quan niệm và ý tưởng, cũng như hiện hữu một cộng đồng cư dân am hiểu. Thư viện số và phần mềm nguồn mở là một sản phẩm tự nhiên của những mô hình trao đổi mở giúp cho xã hội phát triển và thịnh vượng.

Vài phần mềm nguồn mở thư viện số tiêu biểu.

1. Phần mềm xây dựng bộ sưu tập: Greenstone

Greenstone là một trong những hệ thống phần mềm nguồn mở thư viện số nổi tiếng và phổ biến nhất hiện nay. Được phát triển bởi Dự án thư viện số New Zealand của trường đại học Waikato, New Zealand và được UNESCO phân phối, Greenstone là một phần mềm đa ngôn ngữ được dịch sang gần 50 thứ tiếng, bản tiếng Việt do Công ty IES của Phần Lan ở Việt Nam phối hợp với Thư viện Đại học Khoa học

Tự nhiên TP. HCM biên dịch. Phần mềm nguồn mở thư viện số Greenstone và tài liệu hướng dẫn bằng tiếng Việt có thể tải xuống thông qua trang web GreenstoneWiKi (<http://greenstone.sourceforge.net/wiki/index.php/GreenstoneWiki>). Greenstone được dùng để thu gom và biên mục tài liệu theo Dublin Core, đồng thời tổ chức thành bộ sưu tập và xuất ra đĩa CD hay Internet. Greenstone được lưu hành với giấy phép GPL, được dùng rộng rãi khắp nơi trên thế giới như là một “mốt – fashion”. Ở Việt Nam nhiều thư viện đã sử dụng Greenstone để xây dựng nhiều bộ sưu tập có giá trị cao, đáng kể nhất là Thư viện ĐH Khoa học Tự nhiên và Thư viện ĐH Ngân hàng TP. Hồ Chí Minh.

2. Phần mềm gặt hái metadata: dlbox.

dlbox hay “Digital Libraries-in-a-Box” được phân phối bởi Phòng thí nghiệm Thư viện số thuộc trường đại học Virginia Tech, Hoa Kỳ. Phần mềm này được dùng để hỗ trợ việc tích hợp những thành phần mô tả dữ liệu biên mục. Đây là một ứng dụng việc mở rộng giao thức OAI (Open Archives Initiative) cho việc gặt hái metadata (XOAI-PMH). Những thành phần có thể được tải xuống và được xây dựng bởi những công cụ Perl và Java. Nếu chúng ta đã xây dựng cấu trúc cho một bộ sưu tập kỹ thuật số rồi trong một cơ sở dữ liệu quan hệ, thì những thành phần này có thể bổ sung thêm chức năng cho cấu trúc hiện hữu. Gặt hái metadata là để tạo nên những bộ sưu tập mới dựa vào tài nguyên bên ngoài.

3. Phần mềm truy hồi và quản lý thông tin: SiteSearch

Vào Tháng Sáu 2001, OCLC công bố rằng SiteSearch, một bộ gồm những công cụ truy hồi và quản lý thông tin, sẽ không còn là một sản phẩm thương mại nữa. Với việc

phát hành phiên bản 4.0.2a vào 5/2002, SiteSearch được sản xuất và cấp phép với danh nghĩa Giấy phép nguồn mở OCLC và là một trong ít ví dụ trong thế giới thư viện, một sản phẩm thương mại trở thành nguồn mở. SiteSearch có một tầng Z39.50 mạnh và có thể cung cấp một giao diện đối với những bộ sưu tập địa phương hay từ xa. Nhiều máy chuyên đổi và những công cụ khác có thể được dùng để chuyển tải nội dung vào trong SiteSearch, và nó có thể chỉ mục trực tiếp những khổ mẫu MARC, XML, và SGML.

4. Phần mềm chuyển đổi MARC-Dublin Core và Dublin Core-MARC: MarcEdit

MarcEdit được xây dựng bởi Terry Reese, ĐH Oregon, Hoa Kỳ. Phần mềm được thiết kế nhằm hỗ trợ các dạng khác nhau của MARC như MARC 21, UNIMARC, vv... đồng thời cung cấp các chức năng chuyển đổi từ MARC sang Text, Text sang MARC; MARC sang MARC-XML; MARC sang Dublin Core và ngược lại. MarcEdit được thiết kế dựa trên Window COM. Cho phép tận dụng các tiện ích của Windows Scripting Host. MarcEdit là một chương trình lý tưởng cho bất kỳ dự án liên quan đến MARC – từ bảo trì dữ liệu đến tạo biểu ghi cơ sở.

Biên mục trên Web hay tạo lập metadata.

Như chúng ta đã nhận thấy, xây dựng thư viện số đồng nghĩa với việc xây dựng những bộ sưu tập. Cơ sở của bộ sưu tập là những tài liệu do cán bộ thư viện sưu tầm, chọn lọc, và biên mục từng tài liệu. Cũng giống như trong thư viện truyền thống, chúng ta biên mục để tạo nên những phiếu hay biểu ghi mục lục để giúp độc giả tra cứu sách; trong thư viện số với sự hỗ trợ của phần mềm nguồn mở, người ta biên

mục để tạo nên *metadata* cũng nhằm giúp người sử dụng có thể truy cập tài liệu.

Metadata và phiếu mục lục tương đồng với nhau về mục đích sử dụng, thậm chí còn giống nhau về nội dung. Tuy nhiên khác nhau cơ bản là phiếu mục lục thì tách rời với kho sách, trong khi metadata thì thường gắn liền vào nội dung tài liệu. Điều này giải thích rằng, nếu tra cứu vào phiếu mục lục thì ta chỉ được lý lịch của tài liệu mà ta gọi là *biểu ghi thư tịch (bibliographic record)* trong khi tra cứu vào metadata thì ta có thể có được cả nội dung tài liệu đó.

Tài liệu điện tử hay kỹ thuật số ở dạng toàn văn và đa phương tiện phải được đóng gói bằng công nghệ Web. Vì metadata phải gắn liền vào nội dung tài liệu nên cũng phải được đóng gói để trở thành *siêu dữ liệu thư tịch (bibliographic metadata)*. Chẳng hạn như những biểu ghi thư tịch sử dụng khổ mẫu MARC 21 trong thư viện truyền thống, khi sử dụng với tài nguyên điện tử thì phải được đóng gói bằng ngôn ngữ XML để trở thành MARC-XML. Nhiều phần mềm nguồn mở hỗ trợ việc chuyển đổi này, trong đó có MarcEdit.

Nếu sử dụng Dublin Core với 15 thành phần đã được đóng gói bằng XML để biên mục tài liệu truyền thống lần điện tử như Thư viện ĐH Khoa học Tự nhiên TP. HCM đang sử dụng hiện nay thì ta không cần phải chuyển đổi. Nếu có chuyển đổi thì chuyển đổi sang MARC để có thể trao đổi với hệ thống cũ những biểu ghi thư tịch trong thư

viện truyền thống. Cách biên mục này được gọi là biên mục trên Web (Web-based cataloging), cũng đồng nghĩa với việc tạo lập metadata.

Phần mềm nguồn mở, cụ thể là Greenstone hỗ trợ chúng ta biên mục trên web bằng Dublin Core để tạo nên những bộ sưu tập với hệ thống mục lục tra cứu chuyên nghiệp theo tác giả, nhan đề, đặc biệt là tiêu đề đề mục (Subject Heading), vv... mới thực sự là để xây dựng Thư viện số. **Việc sử dụng bộ máy tra cứu Search Engine nhúng vào một cơ sở dữ liệu điện tử để tìm kiếm theo từ khóa một cách máy móc thì chỉ là một hình thức thông tin trên web, hoàn toàn khác hẳn với Thư viện số.**

Kết luận.

Xây dựng bộ sưu tập (collections) là cốt lõi trong việc hình thành Thư viện số. Công việc này đòi hỏi người cán bộ thông tin thư viện tạo lập và gặt hái metadata hơn là chỉ biết sử dụng search engine. Hệ thống nguồn mở là trợ thủ đắc lực cho công việc này.

Sử dụng Hệ thống nguồn mở nói chung và Phần mềm Greenstone do UNESCO quảng bá nói riêng để xây dựng Bộ sưu tập trong việc hình thành Thư viện số là rất cần thiết và là cơ hội đồng đều cho tất cả các thư viện ở Việt Nam hiện nay.

THAM KHẢO

1. *How to build a Digital Library* / Ian H. Witten and David Bainbridge.- New York: Morgan Kaufmann, 2003
2. *Using Open Source Systems for Digital Libraries* / Art Rhyno.- Westport, Connecticut: Libraries Unlimited, 2004
3. *Digital and Digital Libraries Module - Information Management Resource Kit* / FAO and UNESCO, 2005 (CD-ROM)