

## PHẦN MỀM NGUỒN MỞ HÒN ĐÁ XANH VÀ VIỆC PHÁT TRIỂN THƯ VIỆN SỐ

**ThS. ĐOÀN HỒNG NGHĨA**

Head of Software Solutions, 990 Company  
Chief Architect, Integrated e-Solutions Ltd.



*"... Một hòn đá không thể lớn lên được, nó chỉ có thể bị mòn đi hoặc rã ra. Một sinh vật sống có thể lớn lên, thích nghi, hoàn thiện và tiến hoá. Đây cũng chính là sự khác biệt giữa phần mềm thương mại và phần mềm nguồn mở. Phần mềm thương mại không có sự tương tác của người sử dụng, trong khi đó phần mềm nguồn mở được hoàn thiện, thích nghi với yêu cầu của từng người sử dụng và có khả năng tiến hoá theo những nguyên tắc chung của cộng đồng sử dụng để ra ..."*

Bob Young – Kiến trúc sư trưởng RedHat Linux

Tháng 8 năm 2003, Thạc sĩ Nguyễn Minh Hiệp, Giám đốc Thư viện Cao học Trường Đại học Khoa học Tự nhiên có đặt ra vấn đề quản lý các tài nguyên số để phục vụ cho công tác xây dựng thư viện số. Vấn đề quản lý và sử dụng các nguồn tài nguyên số được Thư viện Cao học và Liên hiệp thư viện các trường đại học phía Nam (FESAL) rất quan tâm do nhu cầu cung cấp tài nguyên số cho công tác giảng dạy, học tập và nghiên cứu khoa học. Công ty Integrated e-Solutions Ltd. đã nghiên cứu nhu cầu và khảo sát các phần mềm thương mại và nguồn mở trong nước và trên thế giới. Phần mềm nguồn mở Hòn Đá Xanh (Greenstone) được xác định là giải pháp phần mềm cho bài toán đặt ra.

Hòn Đá Xanh là một bộ phần mềm giúp xây dựng và phân loại các bộ sưu tập (collection) của thư viện số. Hòn Đá Xanh đưa ra một cách tiếp cận mới trong việc tổ chức và xuất bản thông tin trên mạng nội bộ, Internet hoặc trên CD-ROM. Hòn Đá Xanh là kết quả của dự án thư viện số tại trường đại học Waikato, New Zealand (New Zealand Digital Library Project), đã được triển khai và phân phối với sự hợp tác của hai tổ chức UNESCO và Human Info NGO. Hòn Đá Xanh là một phần mềm nguồn mở có sẵn tại địa chỉ <http://www.greenstone.com>, tuân thủ theo GNU General Public License. Mục tiêu của phần mềm thư viện số Hòn Đá Xanh là nhằm giúp cho các tổ chức như các trường đại học, các tổ chức Liên hiệp quốc, các tổ chức phi chính phủ, phi lợi nhuận và các chính phủ trong việc tạo ra các loại thông tin có thể được phân phối trực tuyến hoặc trên các CD-ROM. Phần mềm Hòn Đá Xanh đã được Công ty Integrated e-Solutions Ltd. phối hợp với Thư viện Cao học đã Việt hoá hoàn toàn các thành phần giao diện đến các tài liệu hướng dẫn sử dụng, quản trị hệ thống và phương pháp xây dựng các bộ sưu tập số. Bên cạnh đó Công ty và Thư viện đã biên soạn sách "Thực hành xây dựng thư viện số" và thiết kế Chương trình Tập huấn Xây dựng Thư viện số là những công cụ hết sức hữu ích cho việc xây dựng kho tài nguyên số cho các thư viện.

**T**hư viện số là gì? Thư viện số là một tập hợp các bộ sưu tập số (digital collections) của các đối tượng số hoặc đã được số hoá (digital objects – text, video, audio) với hai khả năng chính:

- Các phương thức truy cập, chọn lọc và hiển thị tài nguyên số (dành cho độc giả)
- Các phương thức xây dựng, tổ chức và lưu hành (dành cho cán bộ thư viện)

Kho tài nguyên thông tin của thư viện số là nơi tập trung các loại tài liệu khác nhau: sách, tạp chí, bài giảng, luận văn, cơ sở dữ liệu, ... dưới nhiều dạng khác nhau: dạng văn bản (text), Postscript, Adobe PDF, Microsoft Word, HTML, CSDL SQL, ... Các tài liệu này đa phần được tổ chức với các cấu trúc bên trong của tài liệu khác nhau (chương, mục, ...) nhưng có thể phân ra từng nhóm (từng bộ sưu tập) theo chủ đề, hoặc lĩnh vực khoa học, ... Khi phân loại các nhóm tài liệu này chúng ta có thể thấy ngay các cấu trúc tổ chức và cách thức biên mục, chỉ mục cần thiết cho các nhóm tài liệu này. Công tác sưu tầm và tổ chức tài liệu hiện nay thực ra là việc tổ chức các sưu tập số, tuy người sử dụng chưa nhận ra điều này. Khác với các tài liệu dạng bản in, việc tin học hoá và số hoá của các thông tin này cho phép chúng ta có khả năng phân loại và đánh dấu (indexing) cấu trúc tổ chức, cấu trúc biên mục tài liệu (biểu ghi thư tịch), các thành phần nội dung (chương, mục, dòng, các hình ảnh, biểu đồ, bản vẽ, cụm từ đặc trưng, ...) Đây là những khả năng mà nhiều phần mềm thương mại hiện nay chưa thể hỗ trợ ví dụ như các phần mềm này không thể cho người dùng tự tổ chức các biểu ghi thư tịch (thường chỉ cung cấp duy nhất một loại biểu ghi) và cách quản lý, đánh dấu nội dung bên trong (thường chỉ đánh dấu các từ). Ví dụ như các sưu tập về âm nhạc, đi kèm theo các bài nhạc có mô tả nốt nhạc hoàn toàn có cấu trúc khác biệt so với một tài liệu là bản đồ chia theo nhiều vùng khác nhau. Sự khác biệt này không chỉ đơn thuần về cách hiển thị mà còn là vấn đề xử lý cấu trúc bên trong của thông tin. Do khả năng cần tích hợp với các dạng thông tin khác nhau, thư viện số cần đáp ứng các nhu cầu lưu trữ tập trung (tại trung tâm tài nguyên) và phân tán (tại các khoa, bộ môn, ...) nhưng vẫn cần đáp ứng nhu cầu quản lý tập trung và tích hợp dễ dàng. Các vấn đề khác không kém phần quan trọng là việc “tự động hoá” công tác nhập dữ liệu (một bộ sưu tập số hoá không thể nhập thủ công), khả năng xuất bản “một” hoặc “vài” bộ sưu tập số từ trong kho tài nguyên ra CD-ROM hoặc DVD-ROM, khả năng “nhập một bộ sưu tập” từ CD/DVD vào trực tiếp trong kho tài nguyên chỉ cần vài thao tác trên giao diện, ... Ngoài ra, vấn đề hỗ trợ đa ngôn ngữ (Việt, Anh, Pháp, Nga, Nhật, Trung Quốc, Ả rập, ...) là vấn đề hết sức cần thiết vì chúng ta không thể phụ thuộc vào phần mềm, chờ đợi các phiên bản mới cho một ngôn ngữ mới nếu như ví dụ chúng ta cần thêm chỉ một bộ sưu tập về bia đá đời Đường của Trung Quốc với các hình ảnh chụp và các chú dẫn tiếng Trung Quốc.

Đây là những câu hỏi mà các phần mềm thương mại khó trả lời được, đơn giản là vì khả năng có được đội ngũ lập trình vừa thông thạo về nghiệp vụ thư viện, các chuẩn quốc tế nhằm trao đổi và xử lý thông tin thư viện, khả năng tích hợp các công cụ hiện đại trên nền nguồn mở. Đây cũng là những yếu kém cơ bản do yếu kém về nghiệp vụ thư viện, về công nghệ thông tin trong thư viện và về năng lực phần mềm của cộng đồng thư viện - công nghệ thông tin của chúng ta.

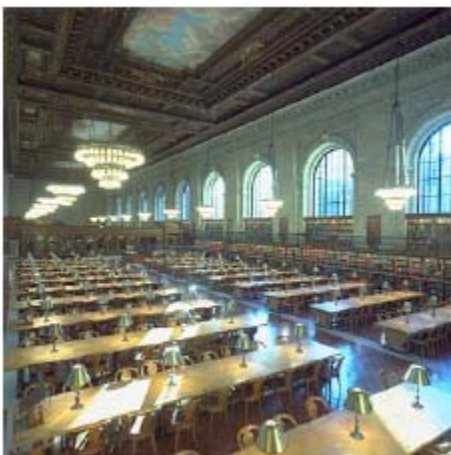
Chúng ta cần làm gì khi xây dựng thư viện số?

Trước khi bắt đầu, bạn nên quan tâm đến những câu hỏi dưới đây:

- Mục tiêu thư viện số của bạn là gì?
- Nhóm đối tượng mà bạn quan tâm?
- Nhóm đối tượng này đa dạng như thế nào, địa phương, tôn giáo hay toàn cầu?
- Số lượng tài liệu bạn muốn có trong thư viện số?
- Tổng cộng bao nhiêu trang?
- Có bao nhiêu tài liệu lại hình ảnh đồ họa?
- Tài liệu có cần thiết được chia thành các phần được tra cứu bởi một số ít người đọc và các phần được tham khảo một cách phổ biến?
- Các tài liệu đã ở sẵn dạng kỹ thuật số chưa?
- Nếu vậy, chúng ở dạng nào?
- Quyền hạn và bản quyền của các tài liệu là gì?
- Ai sở hữu bản quyền?
- Có những tổ chức nào khác có cùng nhóm đối tượng không?
- Bạn có sẵn sàng hợp tác với những tổ chức khác không?
- Ngân quỹ bạn dành cho toàn bộ dự án thư viện số là bao nhiêu?
- Bao nhiêu nhân lực bạn dành cho việc biên tập tài liệu, quét tài liệu và lập trình?
- Cần bao nhiêu máy tính cho dự án?
- Bao nhiêu đĩa CD-ROM bạn muốn phát hành?
- Chúng là miễn phí hay để bán?

Chúng tôi xin đưa ra một so sánh đầu tư (*Global Help Project, Antwerp, UN agencies*)

<b>Đầu tư bản in và phần mềm thương mại:</b>	<b>Đầu tư tài liệu số hoá và Hòn Đá Xanh:</b>
<ul style="list-style-type: none"> <li>○ 160,000 trang, 30,000 ảnh</li> <li>○ 800 sách, 430 tạp chí</li> <li>○ 340 kg</li> <li>○ Máy vi tính, máy chủ</li> <li>○ Phần mềm: 20,000 USD</li> </ul>	<ul style="list-style-type: none"> <li>○ CD-ROM</li> <li>○ Máy vi tính, máy chủ</li> <li>○ Phần mềm: 10,000,000 VND</li> </ul>



Từ Thư viện Quốc hội Hoa Kỳ đến thư viện làng Kataanyi, Kakunyu, Uganda đều sử dụng Greenstone

**H**ành mềm Hòn Đá Xanh là một bộ phần mềm phục vụ công tác xây dựng, lưu hành, truy cập các sưu tập số hoá của thư viện số với đầy đủ các chức năng cơ bản quản lý tài nguyên thư viện thông thường và số hoá:

- Xuất phát từ NZ Digital Library Project, do cộng đồng mã nguồn mở phát triển
- Cung cấp: UNESCO; Human Info NGO; NCSI – Bangalore; UCT – Cape Town, Integrated e-Solutions Ltd. & Thư viện Cao học, ĐHKHTN
- CD-ROM, GNU license
- Giao diện qua Web (Internet Explorer, Netscape) tiếng Anh, Nga, ... và tiếng Việt
- Hoạt động trên tất cả các hệ điều hành thông dụng: Unix/Linux/Windows/Mac OS
- Dạng tài liệu: HTML, Word, PDF, PS, text, email, audio, video, ...
- Các chuẩn dữ liệu: MARC, DC, XML
- Các chuẩn liên thông: Open eBook, SDLIP, Greenstone, OAI, Z39.50
- Tài liệu sử dụng, quản trị, tạo bộ sưu tập bằng tiếng Anh, Nga, Pháp, ... và tiếng Việt
- Phát triển thêm các chức năng theo yêu cầu

Một số ví dụ về các thư viện đang sử dụng phần mềm Hòn Đá Xanh:

The screenshot displays the website for The New York Botanical Garden's LuEsther T. Mertz Library Rare Book Digitization Project. The interface includes a search bar with a 'search' button and a 'titles' button. Below the search bar, there is a search form with dropdown menus for 'text' and 'some' of the words, and a 'Begin Search' button. A section titled 'Searching this collection' provides instructions on how to use the search function. Below this, there is a section titled 'Icons you will encounter when browsing' with several icons and their corresponding actions: a book icon to read the book, a bookshelf icon to look at books on that subject, a document icon to view the document, a folder icon to open the folder and view contents, and a book icon to close the book.



The right side of the screenshot shows a scanned page of text from a book. The page is numbered '12' and titled 'WHITE OAK.' The text describes the distribution of white oaks in the United States, mentioning their presence in the West, Kentucky, Tennessee, and Genesee, as well as their abundance in the middle States, Virginia, and the Southern States. It also mentions their presence in the Alleghany mountains and the banks of the Ohio, within an extent of about 150 miles, reckoning from Brownville, upon the banks of the river Monongahela. The text concludes by stating that the tree is particularly abundant in that part of Virginia and Pennsylvania, situated beyond the Alleghany mountains, and which is comprised between those mountains and the banks of the Ohio, within an extent of about 150 miles, reckoning from Brownville, upon the banks of the river Monongahela. In this space, it was particularly in the environs of Greensburgh, Macónel-Ville, Union-Ville, and Washington C. H., that I saw large forests, nine tenths of which were composed solely of this useful tree, the luxuriant growth of which proves that the

*New York Botanical Garden*



 <p>Project Notes   Browse Collection   Search Collection   Glossary TERMS   NAMES</p> <p>HOME   ILLUSTRATIONS   HELP   CONTACT</p> <p><b>Book Collection Titles (30)</b></p> <p><b>Album of designs of the Phoenix bridge company: successors to Clarke, Reeves &amp; Co., Phoenixville,</b> by Phoenix Bridge Company (Philadelphia: J.B. Lippincott, 1885)</p> <hr/> <p><b>Album of designs of the Phoenixville Bridge-Works,</b> by Clark, Reeves &amp; Co. (Philadelphia: J. B. Lippincott, 1873)</p> <hr/> <p><b>American iron bridges: abstract of the discussion upon a paper submitted by Zerah Colburn,</b> by Zerah Colburn, edited by Charles Manby and James Forrest. (New York: Van Nostrand, 1867)</p>	 <p>Browse Book   Browse Collection   Search Collection   Glossary TERMS   NAMES</p> <p>HOME   ILLUSTRATIONS   HELP   CONTACT</p>  <p><b>Album of designs of the Phoenix Bridge Company</b></p> <p>Corporate author: <b>Phoenix Bridge Company.</b></p> <p>Title: <b>Album of designs of the Phoenix Bridge Company: successors to Clarke, Reeves &amp; Co., Phoenixville bridge works</b></p> <p>Variant title: Phoenix Bridge Company.</p> <p>Publication info: Philadelphia: J.E. Lippincott, 1885.</p> <p>Physical description:</p>
--	---

*Lehigh University Digital Bridges Collections*

 <p>search   titles   uniform titles   genres   dedicate</p> <p>The University of Chicago Library is home to a distinguished collection of over 400 first and early printed editions of musical compositions by Frédéric Chopin. These editions are of particular interest to Chopin scholars and bibliographers because Chopin's works were often published concurrently in several countries with variant texts. Early re-printings were issued using the same plates, causing difficulties identifying first and subsequent impressions. Scholars can establish a sequence of publication by comparing a range of printings. The collection at the University of Chicago, maintained in the <a href="#">Special Collections Research Center</a>, contains evidence that contributes to an understanding of the printing and publishing history of Chopin's works.</p> <p>When completed, the Chopin Early Editions online collection will consist of digitized images of all scores in the University of Chicago Library's Chopin collection. As of February 2003, over a third of the collection is available for viewing online. Users can search or browse the Chopin Early Editions collection via a variety of data points, including titles, genres, plate numbers, publishers and place of publication. Access to individual scores in the digital collection is also available through records in the <a href="#">Library's catalog</a> with links to this Web site. The scanned images of all pages of the scores provide scholars, bibliographers, collectors, dealers, performers, and students with new ways to study this collection.</p>	 <p>HELP   PREFERENCE</p> <p>search   titles   uniform titles   genres   dedicate</p> <p>Adieu à Varsovie : rondeau /</p> <p><a href="#">view score</a>   <a href="#">bibliographic description</a></p> <p><b>Uniform Title:</b> Rondos, piano, op. 1, C minor  <b>Title:</b> Adieu à Varsovie : rondeau /  <b>Title:</b> "Adieu à Varsovie," rondeau in C min. op. 1  <b>Composer:</b> Chopin, Frédéric, 1810-1849.  <b>Composer:</b> Fontana, Julian, 1810-1865.  <b>Publication Info.:</b> London (No. 67, Frith Street, Corner of Soho Square) :Wessel &amp; Co., importers and publishers of foreign music.[ca. 1840]Nouv. éd. corr.  <b>Description:</b> 13 p. of music, 34 cm.  <b>Notes:</b> composé par Frédéric Chopin.      Caption title:      "Edited by I. [i.e. J.] Fontana"--List t.p.      Engraved. Price: 4f.      Also available on the Internet  <b>Subjects:</b> Rondos (Piano)      Piano music.</p>
--	---

*University of Chicago, Chopin Early Editions Collections*

**THAM KHẢO**

1. <http://www.greenstone.org>
2. <http://www.glib.hcmuns.edu.vn>