

MÔ TẢ VÀ SỬ DỤNG TÀI NGUYÊN TRONG CHIẾN LƯỢC PHÁT TRIỂN THƯ VIỆN

ThS. ĐOÀN HỒNG NGHĨA
Integrated e-Solutions Ltd.

"... Có lẽ khó có thể thấy được kết quả lớn khi chúng ta đang ở trong cơn lốc xoáy của sự thay đổi. Nhược điểm thường có của các nhà khoa học và chuyên gia công nghệ là không hiểu được hệ quả của các phát minh của chính mình khi đang trong trạng thái say mê của các khám phá và cải tiến. Đã từ lâu chúng ta bị cuốn hút do lòng khát khao muốn thấu hiểu mọi việc mà không dừng lại để nhận ra rằng sự tiến triển đến những công nghệ mới hơn và hùng mạnh hơn cũng có thể sống cuộc sống riêng của nó..."

Bill Joy – Tác giả của vi (Trình soạn thảo văn bản trong hệ điều hành UNIX)

Nền tảng chính của sự phát triển vượt bậc của loài người là dựa trên sự kế thừa và phát triển liên tục của tri thức. Trong quá trình phát triển này, thư viện là nguồn lưu trữ và cung cấp chính. Sự thành công của châu Âu già nua từ thời trung cổ, ngoài việc tích lũy tư bản còn dựa trên hệ thống đại học lâu đời với các thư viện khổng lồ. Trong những năm gần đây, sự ra đời của máy vi tính và ngành công nghệ thông tin đóng vai trò không nhỏ trong việc phổ biến tri thức, nhất là từ khi Internet được sử dụng rộng rãi. Hệ thống thư viện ngày càng phát triển, hiện đại từng bước từ việc sử dụng phiếu mục lục đến tra cứu biểu ghi qua máy vi tính và đến nay đã có những thư viện số hoàn hảo. Xây dựng chuẩn cho việc mô tả và sử dụng tài nguyên của từng thư viện và việc tích hợp thông tin giữa các thư viện và các hệ thống truy cập và lưu trữ khác, đặc biệt là Internet, trở nên một vấn đề rất đáng quan tâm. Bài viết ngắn này xin đóng góp một số ý kiến trong việc sử dụng các chuẩn mô tả và trao đổi thông tin bên trong và ngoài thư viện.

Hệ thống mục lục (catalogue) là công cụ cơ bản để sử dụng thư viện có hiệu quả. Hệ thống mục lục đóng vai trò vô cùng quan trọng trong việc tìm kiếm, tra cứu sách báo, tạp chí tại các thư viện. Phát triển từ xa xưa, được cải tiến nhiều lần bao gồm các phiếu dùng để tra cứu, phân loại theo các lĩnh vực chuyên môn với các hình thức viết ngắn, viết tắt và các chỉ dẫn tham khảo (cross-reference), các phiếu mục lục trở nên công cụ hữu ích và chuẩn mực trong việc quản lý và sử dụng thư viện.

Kỷ nguyên thông tin (Information Age) cung cấp những công cụ tuyệt vời làm thay đổi hẳn việc sử dụng và quản lý thư viện. Thay cho một chiếc phiếu mục lục 40mmx80mm là vài trăm byte, thay cho những tủ phiếu tra cứu khổng lồ là một hệ thống máy tính gọn nhẹ. Chỉ bằng một số thao tác đơn giản, người đọc có thể tìm thấy ngay tư liệu cần có. Tìm kiếm chỉ theo đề mục và chỉ dẫn tham khảo được mở rộng với việc tìm kiếm theo các từ khoá trong tựa đề và nội dung. Phiếu mục lục thư viện, nay trở thành biểu ghi (record), có thể nối kết

(link) với các thông tin khác, mô tả đầy đủ hơn về tài nguyên (resource) của thư viện. Từ chiếc phiếu mục lục tra cứu thông thường, hệ thống chuẩn Machine Readable Cataloging (MARC) do Thư

Viện Quốc Hội Hoa Kỳ (TVQHHK) tạo nên trong những năm 1960 là một bước tiến lớn, mở rộng khả năng quản lý và sử dụng thư viện lên hàng chục triệu biểu ghi.

Từ một hệ thống cũng nhằm trao đổi thông tin qua thư tín điện tử (email), khai thác tài nguyên có sẵn qua các chuẩn truy cập và trao đổi thông tin dạng tập tin (file) – FTP, Gopher, email, HTTP, World Wide Web – Internet, do các nhà vật lý Trung Tâm Nghiên Cứu Năng Lượng Nguyên Tử Châu Âu (CERN) tạo ra một hướng đi hoàn toàn khác. Sự khác biệt của hai hệ thống chính là tính mở (openness) trong việc trao đổi (exchange) và lưu trữ (storage) tài nguyên. Thư viện điện tử tập trung vào việc chuẩn hoá dạng lưu trữ (storage format) thông tin, tạo ra phương cách duy nhất nhưng thống nhất để truy cập thông tin (information access); trong khi Internet lại chú trọng về tính tương thích và đồng bộ trong trao đổi thông tin và cho phép dạng lưu trữ thông tin mang tính cục bộ.

Tiến trình phát triển của hai hệ thống đều hướng đến việc kết nối với các hệ thống bên ngoài phục vụ việc truy cập của các đối tượng khác và khai thác tài nguyên từ các hệ thống khác. Do chính mục đích chung này nên sự phát triển của mỗi bên đều chú trọng vào việc khắc phục điểm yếu đang có. Các chuyên gia W3C (www.w3.org) đưa ra các chuẩn đóng gói dữ liệu (content encapsulation) để chuẩn hoá dạng thông tin khi chuyển đổi: SGML, HTML, XML, ... Trong khi đó, các thư viện cải tiến các chuẩn

MARC đang có, sử dụng XML và mở rộng các máy chủ để cung cấp và chuyển giao thông tin theo các chuẩn nghi thức trao đổi (protocol) của Internet mà HTTP là một ví dụ.

Sự hội tụ (convergence) tương chừng như đơn giản và tất yếu này, lại chính là đề tài của sự tranh cãi giữa các chuyên gia lĩnh vực thư viện và công nghệ thông tin. Một trong những nguyên nhân cơ bản chính là MARC, chuẩn mô tả tài nguyên thư viện.

Lý do thứ nhất chính là số lượng nguồn thông tin thư viện cung cấp. Theo suy nghĩ trước đây, thư viện cung cấp thông tin từ một nguồn duy nhất. suy nghĩ này hiện nay đang thay đổi dần theo xu hướng khai thác thông tin từ nhiều nguồn khác nhau. Việc lan toả của World Wide Web ngày càng thúc đẩy xu hướng này mạnh mẽ hơn. Có một số thư viện sử dụng song song nguồn tài nguyên dưới dạng MARC và các tài nguyên mới dựa theo chuẩn XML, việc này dẫn đến chi phí đầu tư thay đổi và vận hành lớn, đồng thời làm cho người sử dụng rất lúng túng.

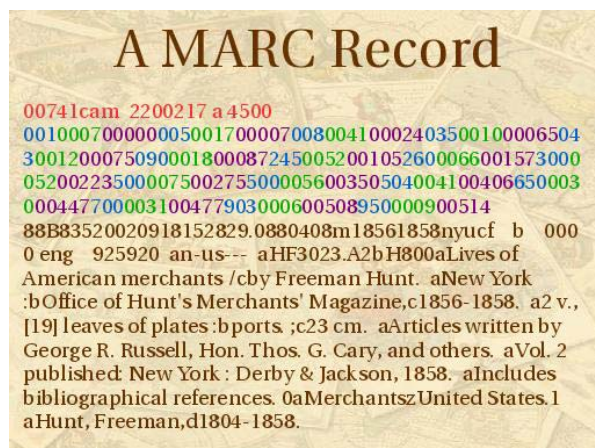
Lý do thứ hai là việc thấu hiểu MARC đòi hỏi các thủ thư mất nhiều năm liên tục tìm hiểu và sử dụng. Việc đào tạo về MARC tại các trường đào tạo thủ thư chỉ cung cấp khái niệm bề mặt.

Ngoài ra việc đào tạo thủ thư còn thiếu đào tạo hoặc cập nhật kiến thức về công nghệ thông tin cũng dẫn đến hạn chế khả năng và tâm lý triển khai công nghệ thông tin mới thay cho công nghệ đang có. Chính tâm lý này cũng là yếu tố cơ bản tại sao một số thư viện sẵn sàng hơn trong việc triển khai ngay hệ thống dựa trên công nghệ cũ.

Lý do thứ ba là việc đầu tư vào các thư viện. Các công ty phần mềm danh tiếng trong lĩnh vực thư viện sẵn sàng vận động (lobby) để bán các phần

mềm dựa trên các công nghệ cũ hoặc xây dựng thêm để tương thích với các công nghệ mới nhưng hoàn toàn theo công nghệ mới để tiết kiệm chi phí. Xây dựng thư viện theo kỹ thuật mới rất tốn kém và theo suy nghĩ hiện nay, thư viện vẫn không phải là bộ phận kinh doanh tạo ra thu nhập cho đơn vị, nên không thể được đầu tư theo nhu cầu.

Lý do thứ tư và cũng là lý do kỹ thuật chính: *MARC được thiết kế quá chuyên dụng cho một mục đích duy nhất.*



Ref: <http://elane.stanford.edu/laneauth/kcs-access2002/img4.html>

Biểu ghi ví dụ trên cho thấy các dữ liệu trong MARC được mô tả rất gọn (compact), tối ưu (chỉ sử dụng đúng số lượng byte cần thiết) và chính xác (mỗi số chỉ mã trường đều có ý nghĩa riêng và cụ thể). Nhưng chính những tính năng này lại là các điểm yếu của MARC. Các trường quản lý (control fields) chứa các thông tin tĩnh và súc tích nhưng không thể thay đổi khi cần giải quyết các vấn đề phát sinh ví dụ như Y2K, các biểu mẫu con, giá trị của trường quản lý khi dạng

biểu mẫu thay đổi ... Việc sử dụng dữ liệu theo phương pháp này là hoàn hảo nếu như hệ thống chỉ có một dạng cơ sở dữ liệu và duy nhất một chương trình truy cập, quản lý và cập nhật. Việc sử dụng lại, kết nối mới đòi hỏi các nhà kết nối, chuyên gia thiết kế lập trình, các thủ thư quản trị tốn nhiều thời gian và công sức. Sự tối ưu quá mức do cải tiến nhiều lần của MARC dẫn đến việc định nghĩa hàng loạt các giá trị và trường ít khi sử dụng (festschrift), giá trị trường sử dụng khác nhau khi ứng dụng (LC subject headings versus gazetteers) và việc truy cập dữ liệu ngày càng khó khăn (hiện nay có hơn 800 trường).

Sự mâu thuẫn trong tổ chức của MARC có thể thấy qua việc sử dụng các giá trị khác nhau cho cùng một ý nghĩa (geographic code versus place of publication code), các dữ liệu chứa tại các vị trí tương tự lại có dạng (format) khác nhau (date), các dữ liệu giống nhau lại lặp lại nhiều lần tại các vị trí khác nhau trong biểu ghi.

Ngôi nhà MARC được trùng tu quá nhiều lần dẫn đến tính mâu thuẫn trong tổ chức.

Dữ liệu tối (dark data) chính là khái niệm để mô tả dữ liệu trong một cơ sở dữ liệu mà chỉ có thể truy cập được qua một giao diện duy nhất. Các hệ thống truy cập và trao đổi thông tin, trong đó có thư viện, hiện nay bao gồm ba phần chính:

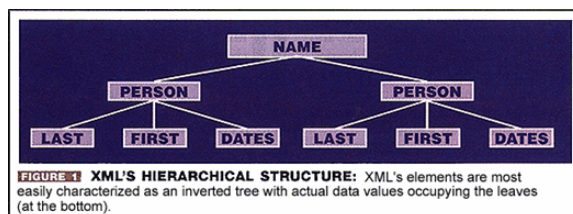
1. Phương thức lưu trữ tài nguyên
2. Chuẩn mô tả khi truy cập và trao đổi
3. Chuẩn nghi thức trao đổi tài nguyên

Các hệ thống thư viện được thiết kế sử dụng MARC hầu như không phân cấp độc lập cho ba phần này, đó là lý do tại sao khó đưa một nguồn tài nguyên thông tin mới vào (vì phương thức lưu trữ tài nguyên không tương thích) hoặc không thể khai thác được (vì không theo chuẩn mô tả nào), hoặc không thể nối kết với hệ thống bên ngoài nào khác (vì không theo chuẩn nghi thức trao đổi tài nguyên nào). Do đó việc mở rộng thư viện về nguồn tài nguyên hay kết nối đều gặp những khó khăn lớn, hầu như đều phải thiết kế lại và nhập lại toàn bộ dữ liệu hoặc phải tạo mới các công cụ chuyển đổi dữ liệu mà hầu hết đều không thể sử dụng trực tuyến (online).

Đây chính là lý do tại sao MARC chỉ trở thành chuẩn dữ liệu của cộng đồng thư viện trong khi đó HTML và người em sinh sau XML lại trở thành công cụ chung hiện nay cho toàn bộ những người sử dụng ứng dụng của công nghệ thông tin hiện đại.

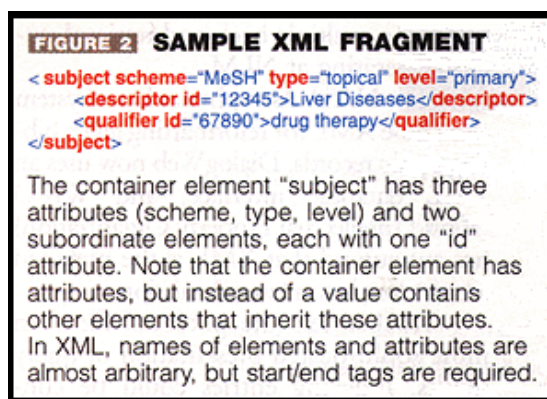
XML mang tính tổ chức chặt chẽ, mỗi văn bản XML đều phải tuân theo các

cấu trúc vật lý và logic chính xác. Về mặt logic, XML chứa các phần tử (element) với hệ thống phân cấp (hierarchy) rõ ràng. Ví dụ sau mô tả hệ thống phân cấp trong một văn bản XML.



Ref: <http://xmlmarc.stanford.edu/LJ/>

Tuy vậy, tính mềm dẻo (flexibility) của XML lại được thể hiện qua chính cấu trúc của nó và cho phép định nghĩa các phần tử dễ dàng và phong phú. Ví dụ sau mô tả việc định nghĩa phần tử với các phần tử con (sub-element) và các thuộc tính (attribute).

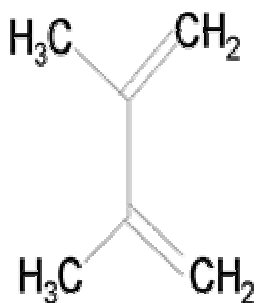


Ref: <http://xmlmarc.stanford.edu/LJ/>

Có nhiều ví dụ cho thấy các cộng đồng khoa học từng bước phát triển ngôn ngữ mô tả của mình dựa trên XML.

MathML chính là ngôn ngữ lưu trữ công thức trên văn bản của các nhà toán học, phát triển dựa trên XML. $\sin^2 x$

Chemistry Markup Language (CML) dùng để mô tả các công thức hoá học. Wireless Markup Language (WML) sử dụng để mô tả trên điện thoại di động.



Ví dụ sau mô tả một phương thức sử dụng XML để định nghĩa “tác giả” với các thông tin phụ nhưng rất hữu ích.

FIGURE 3 INTELLIGENT HYPERLINKING

Author: [Billings, John Shaw, 1838-1913.](#)

- Biographical info
- Portraits
- Web sites
- Works by
- Works about

This hypothetical fragment of a catalog or webliography entry shows how clicking on a single hotlink can yield multiple options for users seeking information about this famous physician/librarian. XLink permits embedded resource locators to create functionality in documents, such as conditional links to scope notes and cross-references in authority records, as well as triggered database searches.

Ref: <http://xmlmarc.stanford.edu/LJ/>

Có hình vì những lý do trên các giải pháp hiện nay được sử dụng khi thiết kế chuẩn dữ liệu metadata cho thư viện có thể chia thành ba nhóm:

- Giải pháp 1: Biên dịch MARC trực tiếp sang XML (MARC XML)
- Giải pháp 2: Tạo mới sử dụng các nguyên tắc của AACR2R (Dublin Core)
- Giải pháp 3: Tạo mới hoàn toàn sử dụng những ưu điểm của XML (MODS)

Ví dụ điển hình nhất của giải pháp 1 là giản đồ (schema) MARC XML của TVQHHK. Giản đồ này thực chất là việc chuyển đổi theo ánh xạ 1-1 giữa các trường của MARC vào các phần tử XML. Bằng phương pháp này, việc trao đổi dữ liệu đã dựa trên chuẩn đóng gói XML (yếu tố thứ hai của hệ thống thư viện) và có thể trao đổi dễ dàng hơn với các hệ thống, tuy nhiên đây là một giải pháp nhiều hạn chế.

Nguyên nhân chính là thay vì *dạng nội dung biểu ghi* tuy đã là XML nhưng *dạng nội dung của từng trường* vẫn không thay đổi. Vì thế việc giải nghĩa (interpretation) các trường dù đã đơn giản hoá (các trường định nghĩa chiều dài không còn là yếu tố quan trọng). Nguyên nhân thứ hai là việc khai thác truy cập dữ liệu trực tiếp (yếu tố thứ nhất của hệ thống) vẫn không có gì thay đổi và vì vậy toàn bộ thông tin vẫn truy cập qua kênh sẵn có. Điều này dẫn đến nhu cầu xây dựng các giao diện với hệ đã có thay vì trực tiếp vào cơ sở dữ liệu, nhưng nhiều hệ thống đang vận hành lại không cho phép xây dựng những giao diện mở này. Việc chuyển đổi này có hạn chế chính vì chưa sử dụng được hết các ưu điểm chính của XML. Tuy vậy, đây là một giải pháp từng phần (partial) đóng góp phương thức chuyển đổi dữ liệu giữa các thư viện đang tồn tại có thể kết nối với nhau. Ngoài ra, ứng dụng của giải pháp này mang lại tính tương thích một cách dễ dàng giữa MARC và MARC

XML. Lợi điểm khác là sự tiết kiệm trong chi phí đào tạo lại cho các thủ thư.

Các giải pháp thuộc nhóm 2 đều dựa trên chuẩn biên mục Anh - Mỹ AACR2R (Anglo-American Cataloging Rules, 2nd Revision). Các giải pháp này đa phần nhằm vào việc tạo ra các tập hợp dữ liệu dùng cho việc định nghĩa các biểu ghi thư tịch (bibliographic record). Mục đích chính là định nghĩa những thành phần thông tin cơ bản (basic information units) của thư viện. Đại diện cho nhóm này là Dublin Core và Metadata Object Description Schema (MODS). Với tính mở cao, nguyên tắc đơn và đã được chuẩn hoá tích hợp với XML, HTML, đây là một trong các giải pháp nhằm mục đích tích hợp các cơ sở dữ liệu MARC và XML với tính trừu tượng cao hơn, và độ tích hợp rộng và sâu hơn đối với các công cụ hệ thống trong công nghệ thông tin. Nguyên nhân chính là việc thay thế định nghĩa trường từ giá trị số sang tên gọi, việc đóng gói cấu trúc trên thể mạnh của XML, đơn giản và dễ hiểu (Dublin Core), đồng thời dùng được các công cụ sẵn có thực hiện việc trao đổi thông tin giữa các thư viện đang dùng MARC với các thư viện hiện đại hoàn toàn không sử dụng MARC (non-MARC).

Resource Description Framework (RDF) là ngôn ngữ metadata mới, phát triển trên XML, mở ra một hướng giải pháp tạo mới hoàn toàn cấu trúc thông tin thư viện theo những nghiên cứu biên mục mới nhất. Hai ví dụ tiêu biểu của giải pháp nhóm 3 là Functional Requirement for Bibliographic Records

(FRBR) và XML Organic Bibliographic Information Schema (XORBIS).

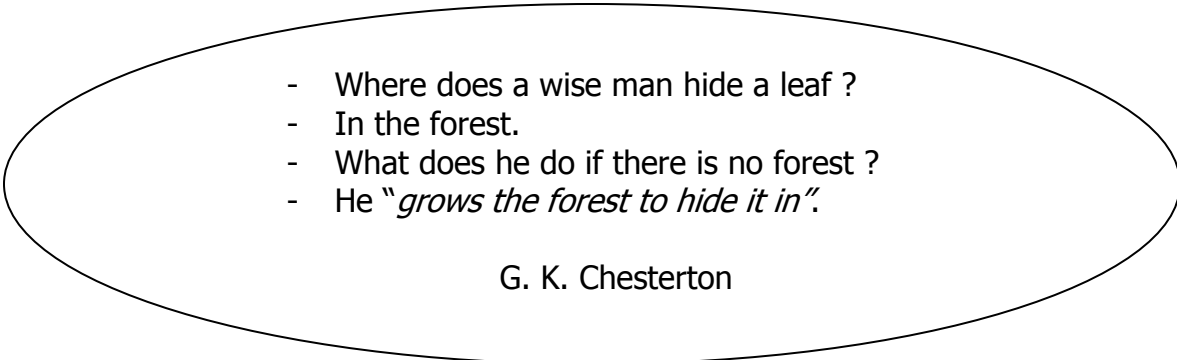
Với những khái niệm mới, mang tính tổng quát, đặc trưng và xây dựng trên nền công nghệ hiện đại trong vòng vài năm trở lại, các giải pháp này về cơ bản đã giải quyết những vấn đề đặt ra để tạo mới hoàn toàn thư viện hiện đại và kết nối với các kho tài nguyên trên Internet. Yếu tố chính của các giải pháp này là:

- Các cấu trúc dữ liệu độc lập, tách rời (modular) và đơn giản đến mức tối thiểu
- Nhấn mạnh tính truy cập (access) cao hơn là mô tả (description)
- Ép cách tham khảo các phần tử mục từ (entry element) theo phương pháp nhất quán
- Sử dụng các cơ sở dữ liệu có thể xác nhận (validate) giá trị cho phép của dữ liệu

Trên đây là một số suy nghĩ tổng hợp dựa trên các tư liệu tổng hợp trong và ngoài nước. Trong tình trạng hiện nay các thư viện Việt Nam ta hầu như chưa có những đầu tư rất lớn như các thư viện nước ngoài và hầu như chưa thống nhất về các chuẩn truy cập, mô tả và trao đổi thông tin. Thiết nghĩ trong việc hiện đại hoá hệ thống thư viện, nên tạo ra những hướng đi và sử dụng giải pháp công nghệ hiện đại và hữu hiệu, tránh việc chuẩn hoá hay Việt hoá những công nghệ đã lỗi thời như MARC, tiết kiệm thời gian và chi phí, thực sự đi tắt đón đầu nhằm đạt được mục tiêu hiện đại hoá và tri thức hoá xã hội đạt đến tầm khu vực và thế giới.

TÀI LIỆU THAM KHẢO

1. Using Dublin Core, Diane Hillmann, 2001-04-12, <http://dublincore.org/documents/2001/04/12/usageguide>
2. Encoding Dublin Core Metadata in HTML, RFC 2731, J. Kunze, 1999-12-30, <http://www.ietf.org/rfc/rfc2731.txt>
3. Beyond the Dublin Core: Rich Metadat and Convenience-of-Use Are Compatible After All, Roger Clarke, 199-07-11, <http://www.anu.edu.au/people/Roger.Clarke/II/DublinCore.html>
4. MARC and XML: A Choice or Replacement, Kevin S. Clarke, <http://elane.stanford.edu/laneauth/kcs-acces2002/img0.html>
5. What is a MARC record and why is it important?, Library of Congress
6. The Dublin Core, simplicity or complexity?, Warwick Cathro, National Library of Australia, <http://nla.gov.au/nla/staffpaper/cathro2.html>
7. Hapy Birthday, XML!, Dave Hollander and C. M. Sperberg-McQueen, <http://www.w3.org/203/02/xml-at-5.html>
8. Academic Annual Meeting 201 of China Society for Library Science, National Library of China, <http://ndl.go.jp/en/publication/cdnla0/042/421.html>
9. XML: Libraries's Strategic Opportunity, Dick R. Miller, <http://xmlmarc.stanford.edu/LJ/>
10. Expressing Simple Dublin Core in RDF/XML, Dave Beckett, Eric Miller and Dan Brickley, 2002-07-13, <http://dublincore.org/documents/2002/07/13/demes-xml/>
11. XML, XSLT, XLink, XHTML, RDF standards, <http://www.w3.org/>
12. Chemistry Markup Language (CML) Tranformation, <http://www.ch.ic.ac.uk/rzepa/chimeral/>
13. Adding Luster to Librarianship XML as an Enabling Technology, <http://elane.stanford.edu/laneauth/Luster.html>
14. Wireless Markup Language (WML), <http://www.wapforum.org/>

- 
- Where does a wise man hide a leaf ?
 - In the forest.
 - What does he do if there is no forest ?
 - He "*grows the forest to hide it in*".

G. K. Chesterton