VẤN ĐỀ SỐ HOÁ KHO TƯ LIỆU HÁN-NÔM THEO CHUẨN DUBLIN CORE TẠI VIỆN THÔNG TIN KHOA HỌC XÃ HỘI

Ts. Ngô Thanh Nhàn, *Đại học Temple* (Hoa Kỳ) &
Nhóm issi_dc *Viện Thông tin Khoa học Xã hội* (Việt Nam)
12 tháng 12 năm 2007

Tóm tắt

Bài này báo cáo đợt khảo sát khả thi việc số hoá kho tư liệu Nôm tại Viện Thông tin Khoa học Xã hội (VTTKHXH) và Trung tâm Triết học, Văn hoá và Xã hội Việt Nam thuộc Đại học Temple (CVST) theo chuẩn Dublin Core nhằm tăng lượng người đọc. Chữ Nôm (gồm cả Hán Việt) là chữ quốc ngữ hệ biểu ý ghi lại tiếng Việt hơn 10 thế kỷ cho đến thập niên 1920. Trong khoảng thời gian dài ấy, mọi tư liệu hành chính, triều đình, bác học, tôn giáo, y dược, văn hoá, sử địa, pháp luật,… và gia đình đều ghi bằng chữ Nôm. Tư liệu di sản văn hoá, như gia phả, vật dụng, đền đài, lăng tẩm, bia mộ,… ghi bằng chữ Nôm vẫn còn đang có nguy cơ bị hư hỏng vĩnh viễn và mất mát kho tri thức ghi bằng thứ chữ này. Số hoá kho tư liệu Nôm gồm có việc quét và chụp ảnh, cùng với việc thiết lập một kho chữ Nôm mã hoá theo chuẩn đa ngữ quốc tế (như Unicode và ISO/IEC 10646). Mục tiêu chính là đưa hệ chữ Nôm vào nền kỹ thuật mạng web quốc tế để thừa hưởng sức mạnh xử lý tri thức của nó.

Kho tư liệu Nôm của VTTKHXH gồm khoảng 20.000 đơn vị sách, bản đồ, bản rập, ảnh, v.v. nằm trong chương trình đưa vào hệ thống thư viện. Hợp tác giữa VTTKHXH và CVST gồm ba phần chính tập trung vào mô hình thư viện mạng internet: số hoá ảnh, số hoá văn bản, và sắp xếp theo chuẩn thư viện quốc tế Dublin Core. Chúng tôi sẽ tiếp tục xem xét các tư liệu Nôm và trình bày trang mạng theo một hay vài cách tìm dùng yếu tố Dublin Core, theo cách truy cập của google.

I. Giới thiệu

Thể theo lời mời của Gs. Hồ Sĩ Quý, Viện trưởng VTTKHXH vào tháng 6, 2006, Gs. Philip Alperson, Gs. Sophie Quinn-Judge, và Ts. Ngô Thanh Nhàn thuộc CVST đã họp và trao đổi với ban lãnh đạo VTTKHXH liên tục từ cuối tháng 7, 2007 nhằm hợp tác nghiên cứu số hoá kho tư liệu Hán-Nôm của VTTKHXH và nghiên cứu tổ chức chương trình đào tạo, huấn nghiệp về khoa học thư viện (*library science*, nay gọi là khoa học thư viện và thông tin, *library and information science, LIS*) với Đại học Temple.

Một chương trình *seminar* diễn ra trong 4 ngày từ 5-8 tháng 3, 2007 tập trung vào việc giới thiệu các tính năng công nghệ thông tin [1] và thông tin thư viện [2,3,4,5] mới nhất trong việc hỗ trợ và khai thác dữ liệu mạng internet. Tiếp đó, một chương trình *workshop* diễn ra trong 5 ngày liên tiếp, từ 4-10 tháng 10, 2007, tập trung vào việc số hoá ảnh, số hoá văn bản, chuẩn và bàn phím nhập chữ Hán Nôm, tổ chức trang mạng công tác, và sơ đồ luồng công tác. Hai cuộc họp dài ngày này, do VTTKHXH và CVST tài trợ, có sự tham gia trình bày của Ts. Ngô Trung Việt (Viện Công nghệ Thông tin Việt Nam) về chuẩn công nghệ thông tin cũng như thông tin thư viện [6], về phông Nôm và chữ Nôm trên mạng internet do nhóm Nôm Na (Hà Nội) [7,8] và Phan Anh Dũng (Trung tâm Công nghệ Thông tin Thừa Thiên–Huế, HueCIT) [9] trình bày.

Hình 1:
Kho tư liệu Nôm, Viện Thông tin Khoa học Xã hội
Ảnh Hoàng Ngọc Sinh, 8.12.2006.

Hình 2: Nhóm *workshop* số hoá issi_dc. Ảnh 10.10.2007.

Các thành viên trong hai buổi họp đã đồng ý thành lập nhóm nghiên cứu sử dụng chuẩn khoa học thông tin Dublin Core—gọi là *issi_dc*—nhằm thích ứng với đặc thù của tư liệu Nôm Việt Nam. Hai kỳ họp đưa ra sơ đồ luồng công tác số hoá kho tư liệu Nôm, bảo vệ bản gốc qua cách sao chụp, phân phối đến nhiều thư viện khác và qua mạng internet. VTTKHXH chưa có phiếu thư mục của kho tư liệu Nôm theo hệ thư viện, và còn trong tình trạng cần được bảo quản chuyên nghiệp.

Bài này tập trung vào phương án phân luồng công tác và những quan sát ban đầu về chuẩn Dublin Core cho kho tư liệu Nôm tại VTTKHXH. Trước hết, chúng tôi xin lược qua đặc trưng của chữ "Nôm" và "quốc ngữ" và quan hệ hữu cơ giữa hai thứ chữ viết, để tránh tình trạng: hoặc chỉ dùng chữ Nôm để điền phiếu thư mục, hoặc chỉ dùng chữ quốc ngữ như thông lệ lâu nay.

2. Vài nét cơ bản về tiếng Việt và hai hệ chữ viết

Tiếng Việt thuộc nhóm Việt Mường trong nhóm Môn-Khơme, ngữ hệ Nam Á [10,11]. Tiếng Việt được coi là thứ tiếng *đơn tiết*, mỗi đơn vị cấu tạo từ và ngữ pháp nhỏ nhất là một âm tiết, thường gọi là một *tiếng* [12]. Tiếng Việt có thanh điệu, mỗi tiếng mang một trong sáu thanh phân bố như sau:

/raŋ/	*Bằng*	*Trắc*	
	trường	*khứ*	*đoản*
cao	rang	rảng	ráng
thấp	ràng	rãng	rạng

Một *tiếng* (hay âm tiết) trong tiếng Việt có ba phần chính: cụm phụ âm đầu, vần và thanh. Trong ví dụ trên, phần chiết đoạn (*segmental*) "rang" (phiên âm /raŋ/) có cụm phụ âm đầu, "r–", và vần, "–ang". Thanh điệu tiếng Việt có tính siêu đoạn (*suprasegmental*). Tiếng cũng là một đơn vị phổ quát của mọi thứ tiếng. Ví dụ, trong tiếng Anh, từ "rang" có một tiếng, trong đó phần chiết đoạn gồm cụm phụ âm đầu "r–", và vần "–ang", phần siêu đoạn của "rang" có thể cao tương đương với thanh ngang.

Tiếng Việt mượn mẫu cấu tạo từ (trừu tượng, khoa học, v.v.) và nhiều tiếng (chữ) của Trung hoa cổ (hay gọi là Hán cổ) và tiếng Trung hoa thời Đường, lúc cả hai chưa có thanh điệu [10,11]. Nay sau hơn 14 thế kỷ, chúng đã được Việt hoá hoàn toàn.

Có **hai thứ chữ viết ghi lại tiếng Việt**: chữ 喃 Nôm (gồm cả Hán Việt) thuộc hệ biểu ý, và chữ quốc ngữ thuộc hệ la-tinh. Chữ Nôm dùng ở Việt Nam khoảng 10 thế kỷ, và được thay thế bằng chữ hệ la-tinh khoảng những năm 1920. Cả hai hệ chữ viết được coi là "quốc ngữ": chữ Nôm là quốc ngữ trước những năm 1920, và chữ hệ la-tinh sau đó. Trong bài này, ta tạm gọi chữ hệ biểu ý là chữ "Nôm", và chữ hệ la-tinh là chữ "quốc ngữ".

Chữ Nôm và chữ quốc ngữ có chung một số đặc điểm cơ bản: chúng phản ánh đơn tiết tính của tiếng Việt bằng cách *ghi mỗi tiếng thành một chữ*. Một chữ trong tiếng Việt gồm có một cụm đơn vị chính tả [13] bao bọc bởi các dấu cách. Chữ quốc ngữ ghi được cấu trúc nhỏ hơn một tiếng nhờ viết theo âm vị của

chữ cái hệ la-tinh [12]. Ví dụ, chữ quốc ngữ cho thấy cấu tạo chi tiết của tiếng, như cụm phụ âm đầu, bán nguyên âm tròn môi /w/ (ghi bằng *o* và *u*); vần gồm hai phần, cụm nguyên âm (nguyên âm dài, nguyên âm ngắn, nhị trùng âm) và cụm chung âm (*coda*); và một trong 6 thanh. Chữ Nôm cũng biểu âm, nhưng gần với ngữ pháp hơn. Ví dụ, chữ quốc ngữ "đá" không phân biệt chữ Nôm 砇 (cho viên/hòn *đá*), hay 跺 (cho cái *đá*), hay 洨 (cho nước *đá* [?]). Những bộ 石 (thạch "đá"), 足 (túc "chân") hay 氵 (băng) bên trái âm 多 "đa", thường lặp lại các loại từ, như "viên/hòn/cục" đứng trước các danh từ chỉ đất đá, loại từ "cái" danh hoá động từ chỉ hành động bằng chân, hay "nước" trước các danh từ thể lỏng, v.v.

Như vậy, ngoài lý do kỹ thuật trong việc phân luồng công tác số hoá, sự có mặt của chữ quốc ngữ song song bên cạnh chữ Nôm trong nguyên bản cho phép các chuyên gia không thông thạo chữ Nôm tham gia công tác số hoá, mà còn giúp mở rộng số người đọc không biết chữ Nôm hay sinh viên Nôm học.

3. Vài nét về phương án phân luồng công tác LIS

Sau khi thành lập ngày 8 tháng 10, 2007, nhóm 20 chuyên gia VTTKHXH và CVST, gọi là **issi_dc**, đề nghị sơ đồ số hoá LIS kho tư liệu Nôm VTTKHXH như trong Hình 3.

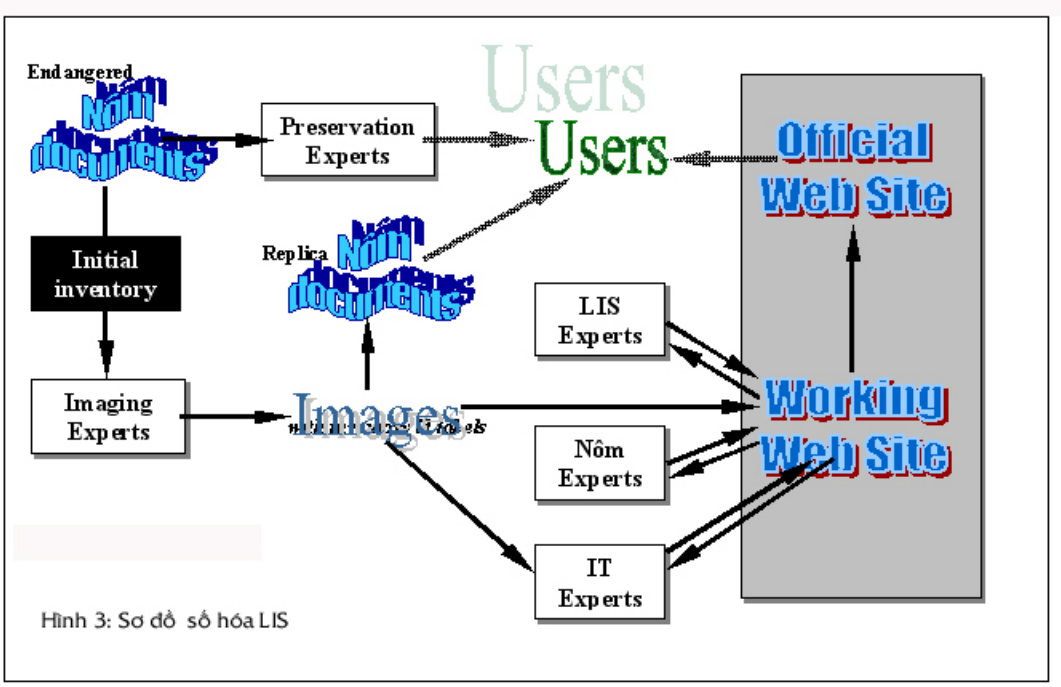
Sơ đồ phân luồng công tác trên có một số đặc điểm sau đây, dựa trên cơ sở tư liệu Nôm hiện tại:

1. **Luồng công tác chụp và xử lý ảnh** theo chuẩn quốc tế, trong đó có microfilm, và ảnh số hoá dùng máy quét hay máy ảnh có độ phân giải cao. Trong luồng công đoạn này có công tác đặt ***mã kho nội bộ*** có hệ thống, không bị trùng, nằm trong trường Mã hiệu (*identifier*) của chuẩn Dublin Core, để có thể theo dõi toàn bộ kho tư liệu trước khi được xử lý bởi các chuyên gia thư viện. Khi có bản sao chụp, mọi công tác, kể cả công tác bảo tồn, có thể xảy ra độc lập tuỳ theo hoạt động của Viện.

2. **Luồng công tác bảo tồn** bản gốc sau khi sao chụp bởi chuyên gia bảo tồn, bảo tàng.

3. **Luồng công tác sao** để phục vụ bạn đọc khắp nơi. Đây là sản phẩm đầu tiên của phương án số hoá. Dụng cụ (máy in màu và đen trắng có độ phân giải cao), phần mềm (xử lý ảnh), vật liệu (giấy dó, kim, chỉ khâu, mực in không acít), và nhân sự (CNTT và đóng sách) dùng trong công đoạn này không tốn kém và có thể thành một hoạt động thường xuyên của Viện, nhằm phổ biến tư liệu cho các thư viện khác phục vụ bạn đọc và nhà nghiên cứu ở mọi nơi.

4. **Luồng công tác số hoá văn bản mạng internet**. Luồng công tác này cần sự hỗ trợ mạnh mẽ của chuyên gia CNTT có hiểu biết về tổ chức số hoá ảnh, hệ phần mềm với cơ sở dữ liệu thư viện, hệ phần mềm bàn phím và phông Hán Nôm, các chuẩn CNTT như XML, MySQL.



Hình 3: Sơ đồ số hóa LIS

5. **Luồng công tác số hoá văn bản Nôm**. Công tác số hoá văn bản Nôm (gồm cả Hán Việt) cần được một vài chuyên gia Nôm hướng dẫn nhóm nhập liệu chữ Nôm và nhóm phiên quốc ngữ.

6. **Luồng công tác thông tin thư viện**. Công tác thông tin thư viện LIS là công tác chính của toàn bộ phương án số hoá, và tham gia hướng dẫn công tác đặt mã kho vào tên gọi các tư liệu đã có mã kho nội bộ để theo dõi các hoạt động sao chép và số hoá văn bản (nhập chữ Hán Nôm và phiên âm quốc ngữ) cũng như hướng dẫn điền mẫu tư liệu Dublin Core.

Ở đây, công tác cần nghiên cứu trước tiên là chuẩn Dublin Core. Phần kế, chúng tôi xin điểm qua những quan sát về tư liệu Nôm trên nền chuẩn thông tin thư viện Dublin Core.

4. Sơ lược về chuẩn thông tin thư viện Dublin Core

Chuẩn Dublin Core – chuẩn cơ bản nhất của các hoạt động thông tin thư viện của nhóm Dublin (thành phố Dublin, tiểu bang Ohio, Mỹ) dưới tên Sáng kiến Siêu dữ liệu Dublin Core (hay DCMI, *Dublin Core Metadata Initiative*)[1] năm 1995 nhằm cải tiến chuẩn tìm kiếm, khám phá mọi nguồn thông tin. Mục tiêu của nhóm là làm thế nào để mô tả tư liệu (*resources*)[2] đủ loại dễ dàng mà ai cũng hiểu được, ít tốn kém, đa ngữ, xuyên ngành, xuyên văn hoá, và dễ tìm. Mục tiêu này hợp với yêu cầu của VTTKHXH đối với kho tư liệu Nôm.

Tập yếu tố Dublin Core được Tổ chức Chuẩn Quốc tế (ISO) chấp thuận ngày 26 tháng 2 năm 2003 mang tựa đề *Thông tin và Tài liệu – Tập Yếu tố*[3] *Siêu dữ liệu Nền Dublin* [14]. Trong phần "Bản quyền", ISO nói đây là "Bản Nháp chuẩn quốc tế". Trong "Lời nói đầu", ISO cho biết chuẩn Dublin mang số hiệu ISO 15636:2003(E), do Tiểu ban 4 (còn gọi là *Tương tác Kỹ thuật*) của Uỷ ban Kỹ thuật 46 (còn gọi là *Thông tin và Tài liệu*). Ngày 22 tháng 5, 2007, chuẩn này được Tổ chức Tiêu chuẩn Thông tin Quốc gia NISO thông qua dưới tên ANSI/NISO Z39.85-2007 [15]. Chuẩn Dublin Core có bản dịch ra chữ Trung hoa giản thể do Thư viện Thượng Hải bảo trì dưới tên *Đô-bá-lâm [Dublin] hạch tâm nguyên số cứ nguyên tố tập* này 28 tháng 8 năm 2006 [16]. Việt Nam chưa có cơ quan nào làm động tác dịch và tham gia hoạt động Dublin Core.

Nhóm công tác issi_dc đề nghị bảng tiếng Việt tương ứng sau khi tham khảo các nghiên cứu khác trong nước [17].

Các chuẩn và bảng từ ngữ chuẩn liên quan đến Dublin Core cần tham khảo và sử dụng khi lập phiếu thư mục, như mã tên quốc gia, mã tên ngôn ngữ 2-chữ số, 3-chữ số và 4-chữ số, thẻ căn cước ngôn ngữ 4-chữ số, mã chữ viết, từ vựng thể loại tham cứu, thể loại lưu trữ internet, tham chiếu tư liệu đồng nhất URI, siêu dữ liệu Dublin Core để truy cập tư liệu, địa danh Getty, mẫu ngày giờ, v.v. [15]

Trong đợt khảo sát sơ khởi này, tương tự như kinh nghiệm xử lý tư liệu cổ Trung, Nhật, Hàn, nhóm nghiên cứu Dublin Core issi_dc đã rà soát sơ khởi thông lệ in ấn truyền thống Việt Nam và sự thích hợp của nó với tập yếu tố Dublin Core. Ví dụ, mỗi dữ kiện ghi bằng chữ Nôm phải có phiên âm quốc ngữ kèm theo. Thời điểm xuất bản thường chỉ vào hoàng đế đương thời và số năm tại vì—đếm từ số 1, thay vì 0, và hoàng đế kế tiếp năm thứ nhất trùng năm với hoàng đế kết thúc trị vì trước đó (xem Phàm lệ của *Đại Việt Sử ký toàn thư*). Hoàng đế thường có quyền nắm toàn bộ ván in, nghĩa là nhà xuất bản có khi cũng là đức vua tại vị, và kinh đô cũng có thể là nơi xuất bản. Cách tính trang tư liệu Nôm bằng tờ, mỗi tờ 2 trang, có "số tờ" in trên cạnh xếp: cạnh xếp in tên sách và số trang. Tư liệu thường không có tên tác giả hay tập thể tác giả, v.v.

---

[1] Thư viện Thượng Hải đã dịch ra tiếng Trung giản thể năm 2006 [16]. Từ "*Core*" dịch là 核心 "hạch tâm". Chúng tôi xin để nguyên chữ "Dublin Core" vì cặp từ này đã thành tên gọi chung. Chúng tôi xin dùng từ "siêu dữ liệu" vì sát với cấu tạo hình vị tiếng Anh (*meta-* "siêu", *data* "dữ liệu"), và cũng sát với nội dung cấu tạo của Dublin Core nhằm tìm ra tên gọi của các dữ liệu có chung một tính chất.

[2] Thư viện Thượng Hải dịch từ "*resource*" là 资源 "tư nguyên". Chúng tôi xin dùng từ "tư liệu" trong bài này thay cho "tư nguyên" hay "tài nguyên" vì chỉ dùng cho kho tư liệu Nôm.

[3] Thư viện Thượng Hải dịch từ "*element*" là 元素 "nguyên tố". Chúng tôi xin dùng từ "yếu tố", và "trường" khi nói đến ứng dụng trong cơ sở dữ liệu.

Bảng 1: Tập 15 yếu tố nền Dublin Core và tiếng Việt tương ứng.

STT	Tên	Tên tiếng Việt	Định nghĩa
1.	Title	*Tựa*	Tên đặt cho tư liệu (Tr.[4] *đề danh*, **题**名).
2.	Creator	*Tác giả*	Thực thể [tên] chính có trách nhiệm chính làm ra tư liệu. (Tr. *sáng kiến giả*, **创**建者).
3.	Subject	*Chủ đề*	Tiêu đề của tư liệu, dùng từ khoá, mã phân loại, hay từ vựng sẵn (khác trường 14. *Phạm vi*), (Tr. *chủ đề/kiện từ*, 主**题**/**键词**).
4.	Description	*Mô tả*	Điều cần nói về tư liệu, ví dụ, tóm tắt, mục lục, hình/ảnh, hay mô tả (Tr. *miêu thuật*, 描述).
5.	Publisher	*Nhà xuất bản*	Thực thể [tên] có trách nhiệm làm cho tư liệu được sử dụng [*available*] (Tr. *xuất bản giả*, 出版者).
6.	Contributor	*Cộng tác*	Thực thể [tên] có trách nhiệm đóng góp vào tư liệu (Tr. *kỳ tha trách nhiệm giả*, 其他**责**任者).
7.	Date	*Thời điểm*	Thời điểm hay thời đoạn dính dáng tới một sự cố trong đời của tư liệu. Đề nghị dùng chuẩn ghi ngày giờ W3C DTF của ISO 8601:1988(E), ví dụ, 2007-12-28T19:20:30.45 +07:00 (+07:00 chỉ múi giờ Hà Nội so với Greenwich) (Tr. *minh kỳ*, 日期).
8.	Type	*Thể loại*	Bản chất hay thể loại của tư liệu. Nên dùng Từ vựng về Thể loại DCMI [*Type Vocabulary*] (khác trường 9. *Dạng thức* mang tính vật lý) (Tr. *tư nguyên loại hình*, **资**源**类**型).
9.	Format	*Dạng thức*	Dạng thức của hồ sơ, dạng vật lý, hay kích thước của tư liệu. Nên dùng từ vựng sẵn trong *Internet Media Type* MIME (Tr. *cách thức*, 格式).
10.	Identifier	*Mã hiệu*	Tham chiếu không mơ hồ của tư liệu trong một bối cảnh nhất định (Tr. *tư nguyên tiêu thức phù*, **资**源**标识**符).
11.	Source	*Nguồn*	Tư liệu [mã hiệu] được sử dụng để thành hình tư liệu này (Tr. *lai nguyên*, 来源).
12.	Language	*Ngôn ngữ*	Ngôn ngữ của tư liệu, dùng chuẩn RFC 4646 (Tr. *ngữ chủng*, **语种**).
13.	Relation	*Tham chiếu*	Tư liệu liên hệ [mã hiệu của tư liệu khác] (Tr. *quan liên*, **关联**).
14.	Coverage	*Phạm vi*	Chủ đề không, thời gian của tư liệu, không gian sử dụng / thẩm quyền pháp lý mà tư liệu có hiệu lực, một ví dụ, danh sách địa danh [*Thesaurus of Geographic Names*] (Tr. *phủ cái phạm vi*, 覆盖范**围**).
15.	Rights	*Quyền*	Các quyền trên hay liên quan đến tư liệu, kể cả tác quyền, sở hữu (Tr. *quyền hạn*, **权**限 [*quản lý*, 管理]).

[4] "Tr." trong cột "Định nghĩa" rút ra từ bản dịch chuẩn DCMI tiếng Trung của Thư viện Thượng Hải.

Tham khảo một vài tư liệu Nôm tại VTTKHXH, có một số điểm làm rõ hơn cách chúng ta định trị và khai triển một số yếu tố:

— Thêm tiểu trường *Quốc ngữ* (gọi là yếu tố chi tiết, *element refinement*)[5] với nghĩa "phiên âm quốc ngữ" của các chữ Nôm trong vào các trường 1. *Tựa*, 2. *Tác giả*, 3. *Chủ đề*, 4. *Mô tả*, 5. *Nhà xuật bản*, 6. *Cộng tác*, 7. *Thời điểm xuất hiện*, và 11. *Nguồn*. Riêng tiếng Việt, Dublin Core cần tiểu trường này, và có thể áp dụng luôn cho những nước có nhiều hơn một thứ chữ viết, ví dụ như tiếng Trung, có cách ghi phanh âm (pinyin), chủ âm quan thoại, hệ Wade, hệ Yale, v.v.

— Thêm 3 tiểu trường *Quyển số*, *Số trang*, và *Kích thước* vào trường 4. *Mô tả*. "Kích thước" theo định nghĩa của DCMI có thể là tiểu trường của trường 9. *Dạng thức*. Tuy nhiên thông lệ mô tả tư liệu, số trang và kích thước thường đi chung.

— Thêm tiểu trường *Mã kho tạm* vào trường 10. *Mã hiệu* để làm sổ kho (*inventory*) sơ khởi.

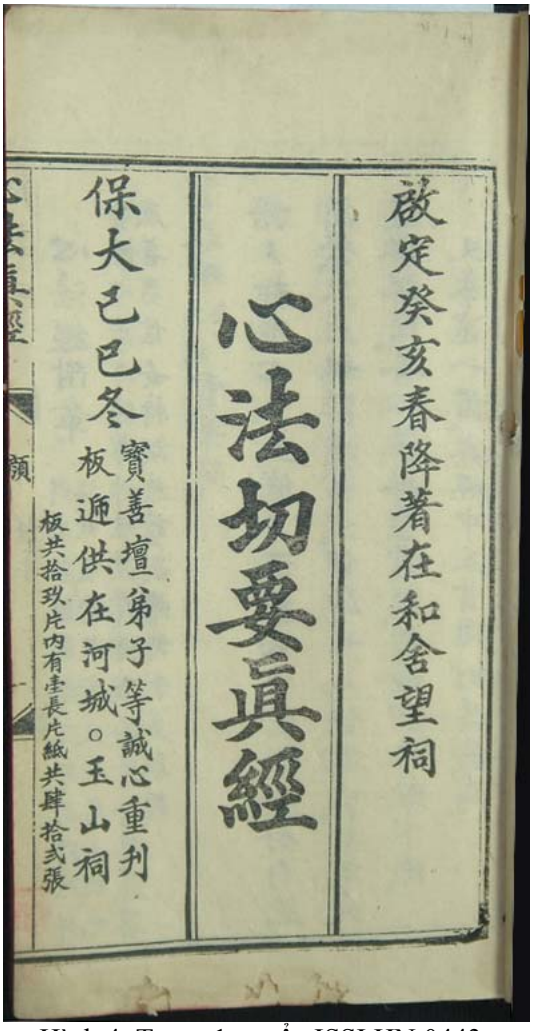
— Xác định trường 12. *Ngôn ngữ*, theo RFC 639-1 (mã 2-chữ số) tiếng Việt là "vi" và RFC 639-2 (mã 3 chữ/số) là "VIE". Quốc ngữ, họ la-tinh, có mã chữ viết "la-VN" hay "vi-Latn". Nhóm issi_dc tạm đặt thêm tiểu trường *Script/Chữ viết* và đặt trị "vi-Nôm" (hoặc "vi-Hani" theo RFC 3066 và ISO 15924, tuy hai chuẩn này chưa nhắc đến chữ Nôm).

— Xác định trường 15. *Quyền* có trị là "Viện Thông tin Khoa học Xã hội".

Vì đã có nhiều cơ quan đang cần sử dụng Dublin Core, Tổng cục Tiêu chuẩn, Đo lường và Chất lượng cần ra một bộ chuẩn Dublin Core tiếng Việt và nộp cho ISO. Chúng tôi tạm gọi là Dublin Core Việt.

Một ví dụ điền mẫu Dublin Core Việt cho tư liệu có mã kho ISSI HN 0443, 心法切要真經 "Tâm pháp thiết yếu chân kinh".

a. **Trang 1**, có ảnh "page001.jpg", xem Hình 4. Trang đọc từ phải sang trái, từ trên xuống dưới, cách cắt câu theo cỡ chữ, dáng chữ, khoảng trống trên và dưới trong cột in, như sau:



Hình 4: Trang 1 quyển ISSI HN 0443.

Cột 1: 啓 定 癸 亥 春 降 著 在 和 舍 望 祠
Khải Định quý hợi xuân giáng trước tại Hoà xá vọng từ
Cột 2: 心 法 切 要 真 經
Tâm pháp thiết yếu chân kinh
Cột 3: Phần chính: 保 大 己 巳 冬
Bảo Đại kỷ tỵ đông
Cột 1 (phải): 寶善壇弟子等誠心重刊
Bảo Thiện Đàn đệ tử đẳng thành tâm trùng san
Cột 2 (trái): 板逅供在河城◦玉山祠
Bản cấu cúng tại Hà thành ◦ Ngọc Sơn từ
Cột nhỏ: 板共拾玖片內有壹長片紙共肆拾弍張
Bản cộng thập cửu phiến nội hữu nhất trường phiến chỉ cộng tứ thập nhị trương.

[5] Khái niệm "*element refinement*" để chỉ một yếu tố nhỏ hơn, rõ ràng và là một thành tố của một yếu tố.

Theo trang trên, sách do thời Bảo Đại in lại ("*trùng san*") năm 1929 của thời Khải Định năm 1923. Sách có 19 bản gỗ, có thêm một bản gỗ dài bên trong, và giấy in ra có 42 tờ.

b. **Trang bìa** và phần liệt kê đầu sách theo Dublin Core, tạm điền như sau:

Bảng 2: Ví dụ điền bản mẫu Dublin Core cho tư liệu ISSI HN 0443

1.	**Title/Tựa**	心法切要真經
	– Quốc ngữ:	Tâm pháp thiết yếu chân kinh
2.	**Creator/Tác giả:**	寶善壇弟子等誠心重刊
	– Quốc ngữ:	Bảo Thiện đàn đệ tử đẳng thành tâm trùng san
3.	**Subject/Chủ đề:**	Buddhism
	– Quốc ngữ:	Đạo Phật
4.	**Description/Mô tả:**	板共拾玖片内有壹長片紙共肆拾弍張
	– Quốc ngữ:	Bản cộng thập cửu phiến nội hữu nhất trường phiến chỉ cộng tứ thập nhị chương
	– Quyển số:	
	– Số trang:	About 500
	– Kích thước:	19 x 26 cm
5.	**Publisher/Nhà xuất bản:**	保大
	– Quốc ngữ:	Bảo Đại
6.	**Contributor/Cộng tác:**	
	– Quốc ngữ:	
7.	**Date/Thời điểm xuất hiện:**	保大己巳冬
	– Quốc ngữ:	Bảo Đại kỷ tỵ đông (1929)
8.	**Type/Thể loại:**	Text [*theo chuẩn DCMI DCT*]
9.	**Format/Dạng thức:**	Book
10.	**Identifier/Mã hiệu:**	
	– Mã kho tạm:	ISSI HN 443
11.	**Source/Nguồn:**	啓定癸亥春降著在和舍望祠
	– Quốc ngữ:	Khải Định quý hợi xuân (1923) giáng trước tại Hoà Xá vọng từ
12.	**Language/Ngôn ngữ:**	VIE [*theo chuẩn ISO 639-2*]
	– Script/Chữ viết:	vi-Nom [*chưa có trong chuẩn quốc tế RFC 3066*]
13.	**Relation/Tham chiếu:**	
14.	**Coverage/Phạm vi:**	Vietnam [*theo chuẩn địa danh TGN*]
15.	**Rights/Quyền:**	Institute of Social Sciences Information

Trị cuả trường 14. *Phạm vi*, TGN [15] đặt là "Vietnam" trong danh sách của có cả "Dai Viet", "North Vietnam" và "South Vietnam". Khi điền mẫu Dublin Core ở trên, phần phiên âm Quốc ngữ là hết sức cần thiết cho người đọc và truy tìm. Hai thứ chữ Nôm và quốc ngữ thật sự hỗ trợ cho nhau về tri thức.

Sau khi xác lập quy trình mạng, nhóm issi_dc dùng mẫu trình bày trang có quy chế tìm theo các yếu tố Dublin Core, có ảnh, có chữ Nôm và phiên âm Quốc ngữ, cũng như có phương án lật trang thuận tiện cho chuyên gia và người đọc. Xem *Hình 5*, trang Dublin Core, và *Hình 6*, trang nội dung cho quyển có mã kho tạm ISSI HN 0443. Hai hình này cho thấy chúng có cùng mẫu dàn trang, và có hai cách để thay đổi nội dung của trang:

a. **Một là thay đổi tư liệu, truy cập theo yếu tố Dublin Core**, ví dụ, xem *Hình 5*, chọn theo Mã hiệu Dublin Core của tư liệu ISSI HN 0443: phương pháp truy cập tương tự như các hệ truy tập tư liệu thư viện mạng hiện nay, và theo phương pháp *googling*, nghĩa là tìm vừa theo các bảng từ vựng chuẩn [15], vừa theo phiếu thư mục (chứa trong thẻ <heading> của tư liệu dạng XML), mục lục, nội dung tư liệu,… có thêm cơ chế hình vị, đồng nghĩa, …

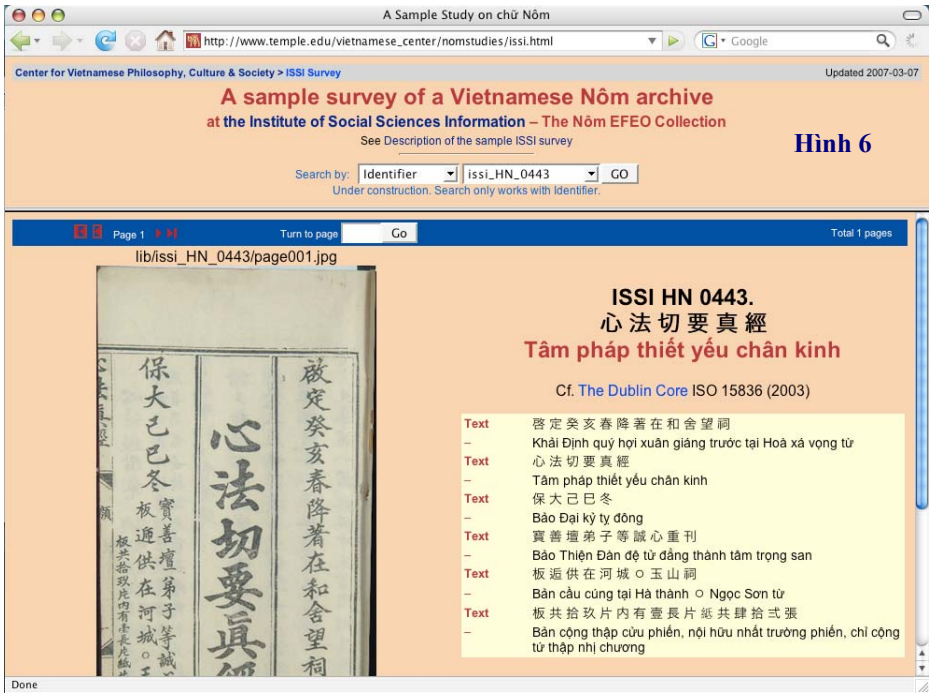
b. **Thay đổi trang** trong một tư liệu đã chọn, như ⏮ "về trang đầu", ◂ "lật trang trước", ▸ "lật trang sau", ⏭ "sang trang cuối" tư liệu, và *Turn to page* ___ "lật đến trang ___" (tự chọn), trong hai *Hình 5* và *6*. Ở đây, chúng tôi dùng một cơ chế trình bày mạng viết bằng PHP:

— thu thông tin ảnh (như page001.jpg trong *Hình 6*, và ảnh của trang có thể lấy xem riêng bằng cách bấm chuột hai lần),

— thông tin thư mục (thẻ thư mục issi_HN_0443.php trong *Hình 6*),

— thông tin chữ Nôm và phiên âm Quốc ngữ (như tệp page001.php trong *Hình 6*), v.v. (như bản dịch, chú thích,…) trong các tệp rời nhau, … để dàn trang, như hai *Hình 5* (trang bìa) và *Hình 6* (lật sang trang 1). Do yêu cầu trên, các công đoạn xử lý ảnh, nhập liệu Nôm/Quốc ngữ, xác định thông tin liệt kê tư liệu Dublin Core, và trình bày mạng có thể tổ chức thực hiện độc lập, không cùng vị trí địa lý (có thể gia công bên ngoài *outsourcing* nếu cần), hoặc phân công trong nội bộ cơ quan.

Hình 5



Hình 6

5. Kết luận

Nhóm nghiên cứu Dublin Core issi_dc dự tính tiếp tục nghiên cứu các văn bản sau:

— ISSI HN 0443 Tâm pháp thiết yếu chân kinh (ảnh chụp được 1 trang)

— ISSI HN 0987 Phù thuỷ thư (ảnh chụp được 58 trang)

— ISSI HN 1011 Tập ghi chép đơn xin nhận ruộng đất canh tác (ảnh chụp được 2 trang)

— ISSI HN 1020 Thái Bình tỉnh, Kiến Xương phủ, Đông Nhuế xã, Đinh tộc Ất chi lập từ (ảnh chụp được 26 trang)

— ISSI TTTS 1261 Thái Bình tỉnh, Kiến Xương phủ, Trực Định huyện, Thuỵ Lũng tổng bộ (ảnh chụp được mới 2 trang)

Các văn bản này đã được đưa lên mạng thao tác theo phân luồng số hoá theo *Hình 3*, chụp ảnh (tiểu ban xử lý ảnh), điền mẫu Dublin Core theo *Bảng 2* (tiểu ban Dublin Core), và nhập chữ Nôm kèm quốc ngữ như *Hình 4* (tiểu ban Hán Nôm) trên trang mạng Đại học Temple (tiểu ban CNTT) tại

www.temple.edu/vietnamese_center/nomstudies/issi.html, dưới sự hướng dẫn của nhóm quản lý.

Tham khảo

1. O'Reilly, T. 2004. *The Open Source Paradigm Shift*. May 2004. On own web page.
2. Antelman, K., Lynema, E. & Pace, A.K. Toward a twenty-first century library catalog, *Information Technology & Libraries* (2006): 128-139.
3. Eberthart, George M. ed. 2006. *The whole library handbook*. Chicago: American Library Association.
4. Martin, Lowell A. 1996. *Organizational structure of libraries*. London: The Scarecrow Press.
5. Schutze, Gertrude. 1972. *Information and library science source book*. Metuchen, NJ: The Scarecrow Press, Inc.
6. Việt, N.T. 2007. *Thư viện số thức*, bộ slides huấn luyện nội bộ, gồm Tập 1, Các vấn đề chung, Tập 2, Chuẩn siêu dữ liệu MODS, METS và Dublin Core, Tập 3: Lưu trữ mở rộng FEDORA và lưu trữ mở OAI. Thông tin riêng.
7. Nhóm Nôm Na. 2004. Quy trình Nôm Na: *Giúp đọc Nôm và Hán Việt* và chữ Nôm trên mạng, *Kỷ yếu Hội nghị Quốc tế về chữ Nôm*. Nxb Văn học, Hà Nội.
8. Nhóm Nôm Na. 2006. 6 phiên bản Truyện Kiều: những vấn đề văn bản học, *Hội nghị Quốc tế về chữ Nôm*, Huế, 6/2006.
9. Dũng, P.A. 2007. *Hệ thống phần mềm và trang web Hán Nôm HueCIT*, bộ slides huấn luyện nội bộ. Trung tâm Thông tin Thừa Thiên–Huế.
10. Haudricourt, A.G. 1954. "De l'origine des tons en viêtnamien," *Journal Asiatique* 242: 68-82.
11. Haudricourt, A.G. 1961. "Bipartition et tripartition des systèmes de tons dans quelques langues d'Extrême Orient," *Bulletin de la Société linguistique de Paris* 56: 163-180.
12. Nhàn, N.T. 1984. *Tiếng và mẫu cấu tạo từ tiếng Việt* [Syllabeme and patterns of word formation in Vietnamese], luận án Tiến sĩ, Đại học New York.
13. Nhàn, N.T. 2001. *Đơn vị chính tả và các đặc điểm của tiếng Việt: chữ quốc ngữ hệ la-tinh, chữ Nôm hệ biểu ý và Unicode/ISO IEC 10646*, Uỷ ban chuẩn Unicode/ISO 10646 [VUIC], 2001.07.01.
14. International Standards Organization. 2003. *Information and documentation — The Dublin Core metadata element set*, ISO/TC 46/SC 4 N515, ISO 15836:2003(E), at http://www.niso.org/international/SC4/n515.pdf, 2003-02-26.
15. National Information Standards Organization. 2007. *Dublin Core Metadata Element Set, Version 1.1*, http://www.dublincore.org/documents/dces/. 2007-05-26.
— *Mã tên ngôn ngữ* 3-chữ số ISO 639-2:1998. <www.loc.gov/standards/iso639-2/langhome.html>
— *Mã tên quốc gia* ISO 3166 <www.din.de/gremien/nas/nabd/iso3166ma/>
— *Thẻ căn cước ngôn ngữ*, Internet RFC 3066. <www.ietf.org/rfc/rfc3066.txt>
— *Mã chữ viết*, ISO 15924. <www.unicode.org/iso15924/codelists.html>
— *Thể loại lưu trữ internet* (Internet Media Types). <www.isi.edu/in-notes/iana/assignments/media-types/media-types>
— *Từ vựng Thể loại tham cứu DCMI*. Đề nghị của DCMI, 2000-07-11. <http://dublincore.org/documents/dcmi-type-vocabulary/>
— *Tham chiếu tư liệu đồng nhất* (Uniform Resource Identifiers, URI): Generic Syntax, Internet RFC 2396. <www.ietf.org/rfc/rfc2396.txt>
— *Siêu dữ liệu Dublin Core để truy cập tư liệu* (Dublin Core Metadata for Resource Discovery). Internet RFC 2413. <www.ietf.org/rfc/rfc2413.txt>
— *Địa danh Getty* <www.getty.edu/research/tools/vocabulary/tgn>
— *Mẫu ngày giờ*, ghi chú của W3C. <www.w3.org/TR/NOTE-datetime>.
16. 上海图书馆 Thượng Hải đồ thư quán. 2006. 都柏林核心元数据元素集1. 1版： 参考描述 [Đô-bá-lâm hạch tâm nguyên số cứ nguyên tố tập 1.1 bản: Tham khảo miêu thuật], http://dc.library.sh.cn/1-1.htm. 2006-08-28.
17. Hội, Nguyễn V. 2007. *Tiêu chuẩn mô tả nguồn tin điện tử trên mạng Dublin Core*. Phòng Báo Tạp chí, Viện Thông tin Khoa học Xã hội. Thông tin riêng.