INITIAL REPORT ON DIGITIZATION
UNDER THE DUBLIN CORE OF THE HÁN-NÔM ARCHIVE
AT THE INSTITUTE OF SOCIAL SCIENCE INFORMATION

Ngô Thanh Nhàn, *Center for Vietnamese Philosophy, Culture & Society, Temple University* (US) &
issi_dc Working Group, *Institute of Social Science Information* (Việt Nam)
December 12, 2007

Abstract

This is an initial report of two surveys on the feasibility of digitization of the Nôm archive, under Dublin Core standards, at the Institute of Social Science Information (ISSI) conducted in cooperation with the Center for Vietnamese Philosophy, Culture & Society of Temple University (CVST). The aim of the digitization is to increase readership. The ideographic Nôm script (Hán Việt included) was the national script in Vietnam for more than 10 centuries until the 1920's. During that period, all administrative, royal, scholarly, religious, medical, cultural, geographical, historical and legal documents were written or printed in Nôm. Vietnam's cultural heritage embodied in family records, palaces, temples, household artifacts, tombstones, steles, etc. containing Nôm texts are in danger of permanent damage and loss, along with the knowledge therein. Digitization of the ISSI Nôm archive, including scanning/photographing, and building associated web texts under multilingual computer character standards (i.e. Unicode and ISO/IEC 10646), is intended to bring the Nôm script onto the web platform to to take advantage of its data mining and processing power, such as googling.

The ISSI Nôm archive contains about 20,000 items—books, maps, and imprints—already part of a library program to increase readership. The cooperation between the ISSI and CVST consists of three main components of an internet library model—digitized images, digitized texts, and listing—under the Dublin Core, as an international library information standard. We will continue to examine the Nôm resources, their suitability for web presentation and their ease of access using Dublin Core elements.

I. Introduction

Responding to the initiative of Prof. Hồ Sĩ Quý, Director of ISSI, in June, 2006, Prof. Philip Alperson, Prof. Sophie Quinn-Judge and Dr. Ngô Thanh Nhàn of CVST have had several meetings with the ISSI leadership since the end of July, 2007 to discuss digitization of the ISSI Nôm collection and training for the ISSI staff in library information science (LIS) in cooperation with Temple University.

A 4-day *seminar* was organized on March 5-8, 2007, as an executive introduction to the most up-to-date information and library technologies [1,2,3,4,5] for

Fig. 1: Nôm Collection, Institute of Social Science Information
Picture by Hoàng Ngọc Sinh, December 8, 2006.

the support of and access to web data. A subsequent 5-day *workshop* was organized on October 4-10, 2007, on image digitization, text digitization, standards and keyboard entries for Nôm, workflow web design, and a digitization workflow chart. These two workshops were organized and financed by the ISSI and CVST. Dr. Ngô Trung Việt (Institute of Informatics, Việt Nam) was invited to lecture on IT standards and LIS [6], the Nôm Na Group (Hà Nội) on Nôm fonts [7,8], and Phan Anh Dũng (Thừa Thiên–Huế Center for Information Technology, HueCIT) on Nôm data entry for the internet [9].
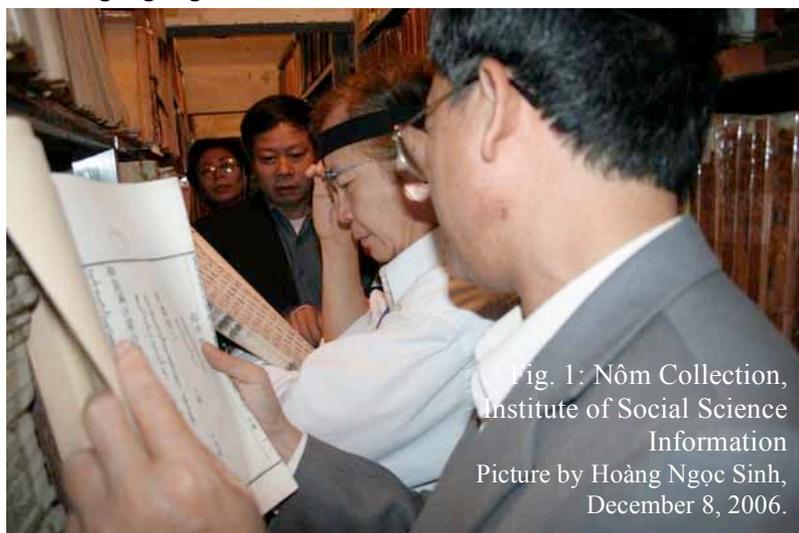
Fig. 2: Participants at the workshop, Oct. 10, 2007.

Twenty participants in the workshops volunteered to form the **issi_dc working group** on October 8, 2008 to study Vietnamese Nôm resources in the light of the Dublin Core standard. The meetings produced a draft digitization workflow for the ISSI Nôm archive, to protect the originals by scanning or photographing, serving more readers via the internet and by sending replica copies to other libraries. It is a complex task to organize a library catalogue for the Nôm archive, which is in dire need of professional preservation.

This paper describes a series of initial observations on the Dublin Core applied to the ISSI Nôm resources. In the next section, we will present the characteristics of the two Vietnamese writing systems, ideographic "Nôm" and latinized "quốc ngữ". We hope to explain the use of both scripts to fill the Dublin Core catalogue card.

2. Basic characteristics of Vietnamese and its two writing systems

Vietnamese belongs to the Việt-Mường subgroup of the Môn-Khmer branch of Austroasiatic languages [10,11]. Vietnamese is *monosyllabic—*the syllable is the smallest morphological and syntactic unit [12]. Vietnamese has 6 tones, each syllable carries one phonemic tone:

/raŋ/	Level	Unlevel	
	long	complex	short
high	rang	rảng	ráng
low	ràng	rãng	rạng

The syllable is a universal unit of speech. A syllable in Vietnamese, *tiếng*, has three main parts: initial consonant, rhyme and tone. The rhyme, *vần*, consists of the vocalic neucleus and the coda. Thus, the segmental syllable for "rang" has the initial consonant *r–* and the rhyme *–ang*. Tone in Vietnamese is *suprasegmental*. Likewise, in English, "rang" /ɹæŋ/ is monosyllabic, having the segmental initial consonant *r–*, the rhyme *–ang,* and no phonemic tone, but with a pitch almost as high as the Vietnamese level tone *ngang*.

Vietnamese has a Sino-Vietnamese (SV) formation pattern for abstract and scientific words, and a series of syllables borrowed from old Chinese and from Chinese of the 唐 T'ang Dynasty, at a time when neither language had tones [10,11]. Today, after more then 14 centuries, they have been totally Vietnamized. The ideograms that represent SV syllables are called Hán Việt.

Vietnamese has **two writing systems**: ideograhic 喃 Nôm (Hán Việt included) and latinized quốc ngữ. Nôm was used in Vietnam for more than 10 centuries, and was replaced by the latinized quốc ngữ in the 1920's. Both writing systems are national scripts: Nôm was the national script before the 1920's. In this paper, we use the term "Nôm" to mean the national ideographic script, and "quốc ngữ", the current national latinized script.

Nôm and quốc ngữ share several characteristics: they reflect the monosyllabicity of Vietnamese by writing each syllable separately—one *tiếng* [syllable] *is written by one chữ*. A *chữ* in Vietnamese consists of

a cluster of orthographic units [13] surrounded by delimiters. Quốc ngữ, thanks to the latin alphabet, records the syllable phonemically [12]. A *chữ* in quốc ngữ shows the phonemic structure of the corresponding *tiếng*, such as members of the initial consonant cluster, rounded semivowel /w/ (written as *o* and *u*), the vocalic cluster (long, short vowels and diphthongs) and the coda cluster of the rhyme, and a tone mark. Nôm is also phonetically based, but its internal composition is closer to the syntactic structure of Vietnamese. For example, "đá" in quốc ngữ does not differentiate between "rock" (Nôm 砙), or "kick" (Nôm 踲), or "ice" (Nôm 氹). The "radical" graphic element—石 *thạch* ("rock"), 足 *túc* ("foot") or 冫 *băng* ("ice")—on the left hand side of the "sound" graph 多 "đa" of the above ideograms, agrees with their syntactic classifiers, such as "viên/hòn/cục" (before nouns of rock-type objects), "cái" (before verbs of action by foot), or "nước" (before nouns of liquid-type objects), respectively.

Thus, besides bringing Nôm to the majority of librarians and readers who can only read quốc ngữ, the presence of quốc ngữ as an element refinement of Dublin Core elements in Nôm also helps readers everywhere inside and outside the country.

3. A sketch of the LIS digitization workflow

On October 10, 2007, the issi_dc working group proposed an LIS digitization workflow for the ISSI Nôm archive, shown in Fig. 3. The draft workflow has several subtasks, based on the current state of the ISSI archive:

1. **Imaging (scanning, photographing and image processing)** of original Nôm resources following the international standard requirements, including microfilming, high resolution hand-scanning, or photographing. This task adheres closely to the *internal inventory coding system* assigned by the ISSI LIS experts to be used in the Dublin Core *Identifier*. After images are produced, all of the other four subtasks can proceed independently.

2. **Preservation** of the originals after imaging by preservation experts: originals can be preserved and stored independently whenever experts and funds are available.

3. **Replication** from the images in hard copy to serve readers in different locations. This is the first product of digitization. Equipment (high resolution color and black-and-white printers), processing software, materials (acid-free *dó* mulberry paper, needle, thread, inks,…) and staff (IT, printing, bookbinding) are inexpensive and ISSI can work to regularly provide replica copies to other libraries and researchers.

4. **Web image processing** requires expertise of IT web experts with software and LIS database management system, XML, etc.
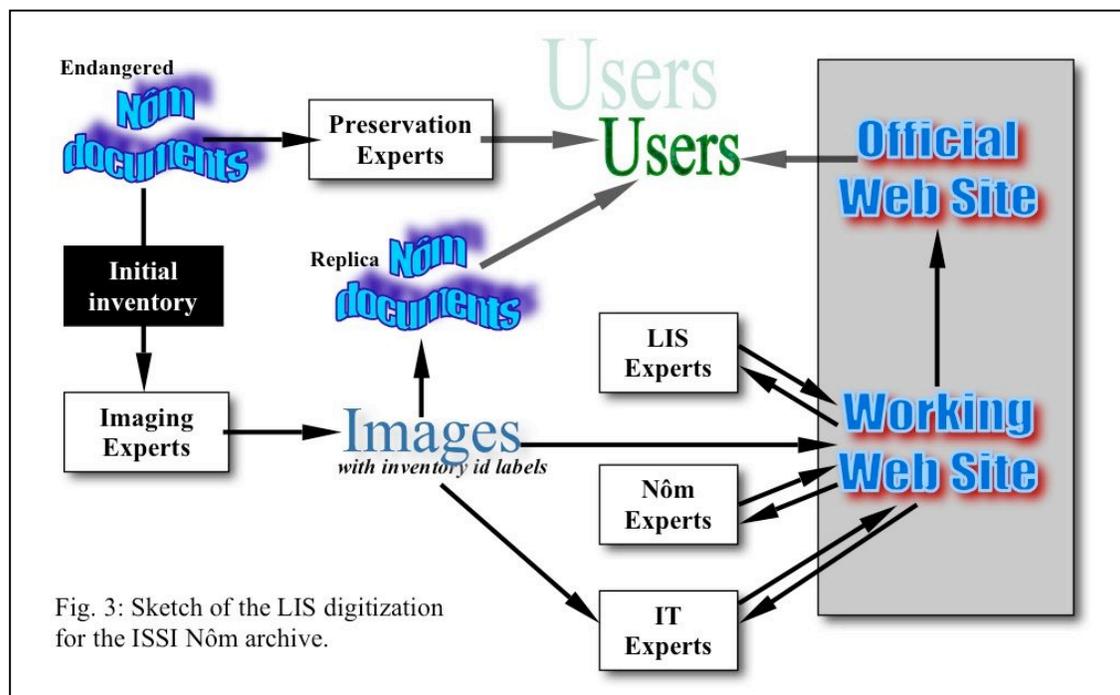

Fig. 3: Sketch of the LIS digitization for the ISSI Nôm archive.

5. **Nôm text digitization** requires expertise in Nôm to assure accurate Nôm text data entry and transliteration from Nôm to quốc ngữ.

6. **LIS data entry** is the key task behind the LIS digitization plan, directing the initial internal inventory coding convention, which affects the naming convention for the entire group of digitization activities, and for searching.

At this point, we turn to the first investigation of the Dublin Core itself, which does not yet have a translation into Vietnamese.

4. A sketch of the Dublin Core in Vietnamese

The Dublin Core—a cross-domain information resource[1] description—is a broad and generic standard to describe a wide range of resources, initiated by the Dublin Core Metadata Initiative (DCMI, Dublin, Ohio)[2] in 1995. The goal is to describe different types of resources in a way that is easy to understand, inexpensive, multilingual and multidisciplinary, cross-cultural as well as easily accessible. This goal coincides with the needs of the ISSI with respect to its Nôm archives.

The Dublin Core element set was adopted by the International Standards Organization on February 26, 2003 under the title *Information and Documentation – Dublin Core Metadata Element[3] Set* [14]. In the "Copyrights" page, ISO states that it is "a Draft International Standard". In the "Foreword", Dublin Core was labelled ISO 15636:2003(E) by ISO Subcommittee 4 (otherwise called *Technical*

Interoperability) of the Technical Committee 46 (otherwise called *Information and Documentation*). On May 22, 2007, the National Information Standard Organization (NISO) adopted this standard under the code ANSI/NISO Z39.85-2007 [15]. This version was also translated into simplified Chinese on August 28, 2006, and maintained by the Shanghai Library [16]. There is no official Vietnamese translation of the Dublin Core at present.

The issi_dc working group proposed a Vietnamese version of the Dublin Core after referring to different attempts at translation in Vietnam [17].

The standards and controlled vocabularies related to the Dublin Core were referenced and used to fill catalogue cards, such as the country code, the 2-digit, 3-digit and 4-digit language identification codes, script code, reference format and storage controlled vocabularies, uniform resource identifier URI, Dublin Core metadata to access resources, Getty thesaurus of geographic names, date and time formats, etc. [15]

In the initial surveys, similar to the experience with China, Korea and Japan in dealing with their ancient resources written in ideographic scripts, the issi_dc working group reviewed the traditional printing and script production in Vietnam as well as the ISSI Nôm archive for the Dublin Core. For example, each data set in Nôm must be accompanied by a transliteration into quốc ngữ. The date usually points to the emperor in power and the year of his reign—starting at 1, not 0; consequently, the year of the outgoing emperor often overlaps the year of the incoming emperor (see 纂修凡例 *Soán tu phàm lệ* [Compilation Conventions] of 大越史記全書 *Đại Việt Sử ký toàn thư* [Complete History of Đại Việt]). The emperor usually owned all print blocks, i.e. s/he might also be the publisher, and the kingdom, the publishing location. The book resources are usually counted by sheets, not pages. Each sheet has two pages, with the name of the book and the sheet number printed on the folded edge. Resources usually do not have name(s) of author(s). And so on.

---

[1] Shanghai Library translates the word "*resource*" to 资源 in Chinese, SV *tư nguyên*. We would like to use the word "*tư liệu*" rather than "*tư nguyên*" or the general term, "*tài nguyên*".

[2] Shanghai Library maintains a version of the Dublin Core in simplified Chinese since 2006 [16]. The word "*Core*" is translated to Chinese as 核心. We shall keep the word "Dublin Core" intact without translating it into Vietnamese. We will use the word "siêu dữ liệu" to be close to the morphological construction of the word "metadata" (*meta-* "siêu", *data* "dữ liệu").

[3] Shanghai Library translates the term "*element*" as 元素 [SV *nguyên tố*]. We will use the word "*yếu tố*" for "element," and "*trường*" when we discuss its presence in a database management system.

Table 1: 15 Dublin Core elements and the corresponding Vietnamese names.

No.	Name	Vietnamese Name	Description in Vietnamese
1.	Title	*Tựa*	Tên đặt cho tư liệu (Ch.[4] *đề danh*, 题名).
2.	Creator	*Tác giả*	Thực thể [tên] chính có trách nhiệm chính làm ra tư liệu. (Ch. 创建者 *sáng kiến giả*).
3.	Subject	*Chủ đề*	Tiêu đề của tư liệu, dùng từ khoá, mã phân loại, hay từ vựng sẵn (khác trường 14. *Phạm vi*), (Ch. 主题/键词 *chủ đề/kiện từ*).
4.	Description	*Mô tả*	Điều cần nói về tư liệu, ví dụ, tóm tắt, mục lục, hình/ảnh, hay mô tả (Ch. 描述 *miêu thuật*).
5.	Publisher	*Nhà xuất bản*	Thực thể [tên] có trách nhiệm làm cho tư liệu được sử dụng [*available*] (Ch. 出版者 *xuất bản giả*).
6.	Contributor	*Cộng tác*	Thực thể [tên] có trách nhiệm đóng góp vào tư liệu (Ch. 其他责任者 *kỳ tha trách nhiệm giả*).
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) (Ch. 日期 *minh kỳ*).
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) (Ch. 资源类型 *tư nguyên loại hình*).
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 (Ch. 格式 *cách thức*).
10.	Identifier	*Mã hiệu*	Tham chiếu không mơ hồ của tư liệu trong một bối cảnh nhất định (Ch. 资源标识符 *tư nguyên tiêu thức phù*).
11.	Source	*Nguồn*	Tư liệu [mã hiệu] được sử dụng để thành hình tư liệu này (Ch. 来源 *lai nguyên*).
12.	Language	*Ngôn ngữ*	Ngôn ngữ của tư liệu, dùng chuẩn RFC 4646 (Ch. 语种 *ngữ chủng*).
13.	Relation	*Tham chiếu*	Tư liệu liên hệ [mã hiệu của tư liệu khác] (Ch. 关联 *quan liên*).
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*] (Ch. 覆盖范围 *phủ cái phạm vi*).
15.	Rights	*Quyền*	Các quyền trên hay liên quan đến tư liệu, kể cả tác quyền, sở hữu (Ch. 权限 *quyền hạn* [管理 *quản lý*]).

---

4   "Ch." in the "Definition in Vietnamese" column was taken from the standard DCMI translation into simplified Chinese by the Shanghai Library.

Several issues become clearer as to how one defines Dublin Core element refinements and how one assigns values to the elements from a sample of the ISSI documents examined:

— To add refinement[5] *Quốc ngữ,* meaning "transliteration into quốc ngữ" side by side with Nôm values in elements 1. *Title*, 2. *Author*, 3. *Subject*, 4. *Description*, 5. *Publisher*, 6. *Contributor*, 7. *Date*, and 11. *Source*, in Vietnamese. This refinement may also be needed by other languages which have more than one script, e.g. Chinese with pinyin, mandarin, the Wade and Yale systems, etc.

— To add three refinements, Volume No./*Quyển số*, Number of Pages/*Số trang*, and Size/*Kích thước* to element 4. *Description*. *Size* according to DCMI can be in element 9. *Format*. However, information such as number of pages and size are usual physical descriptions of resources.

— To add refinement "Inventory Number" *Mã kho tạm* to element 10. *Identifier*.

— To assign value "vi" (according to RFC 639-1 for Vietnamese), and "VIE" (RFC 639-2), to element 12. *Language*. The issi_dc working group temporarily adds refinement *Script/Chữ viết* with value "vi-Nôm" (or "vi-Hani" according to RFC 3066 and ISO 15924), although Nôm has not been found in any standard. The latinized quốc ngữ has script code "vi-Latn".

— To assign value "Institute of Social Science Information" to element 15. *Rights*.

Many institutions in Vietnam have begun to use Dublin Core for their archives. We believe that the General Office of Standards, Metrology and Quality is the appropriate institution to maintain a Vietnamese version of Dublin Core for the ISO.

Below is one example of a Dublin Core page for an inventory code ISSI HN 0443 of the book 心法切要真經 "Tâm pháp thiết yếu chân kinh".

a. **Page 1**, having an image "page001.jpg", see Fig. 4. Reading from right to left, with phrase breaks identified by font changes, font size changes, and blanks on top and bottom of the print columns:
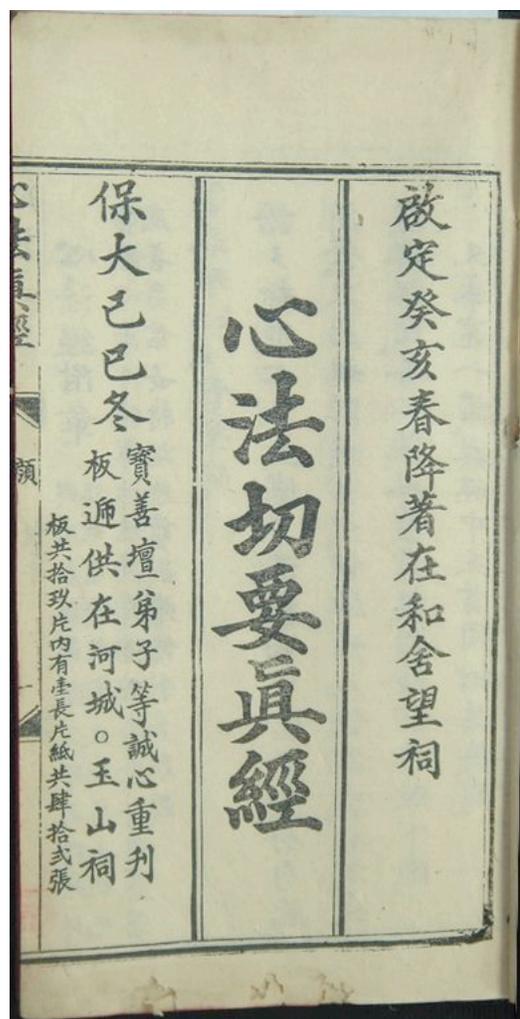

Fig. 4: Page 1 of ISSI HN 0443.

Column 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

Column 2: 心 法 切 要 真 經
Tâm pháp thiết yếu chân kinh

Column 3: Top: 保 大 己 巳 冬
Bảo Đại kỷ ty đông

Column 3.1: 寶善壇弟子等誠心重刊
Bảo Thiện Đàn đệ tử đẳng thành tâm trùng san

Column 3.2: 板迍供在河城。玉山祠
Bản cấu cúng tại Hà thành ∘ Ngọc Sơn từ

Column 3.3: 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 chương.

_____

5 The concept "*element refinement*" is meant to indicate an unambiguous subelement, an elaboration of an element.

The book was reprinted (重刊 *trùng san*) during the reign of Emperor Bảo Đại in the year of the Snake [1929] from the original woodcut carved during the reign of Emperor Khải Định in the year of the Pig [1923]. The book has 19 woodblocks, with one long woodblock inside, and is printed in 42 chapters on paper.

b. **The Cover page** and a catalogue page according to Dublin Core can be filled as follows:

	Table 2: An example for a Dublin Core element set of resource 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式張
	– 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 ty đông (1929)
8.	**Type/Thể loại:**	Text [*cf. 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 [*cf. ISO 639-2*]
	– Script/Chữ viết:	vi-Nom [*not yet in RFC 3066*]
13.	**Relation/Tham chiếu:**	
14.	**Coverage/Phạm vi:**	Vietnam [*cf. not yet in TGN*]
15.	**Rights/Quyền:**	Institute of Social Sciences Information

We set the value of element 14. *Coverage* to "Vietnam", contrary to those in TGN [15] specifying only "Dai Viet", "North Vietnam" and "South Vietnam". The element refinement *Quốc ngữ* transliteration is necessary for readers and resource searchers. Ideographic Nôm and latinized quốc ngữ, thus, complement each other.

After establishing the web process for the Nôm digitization workflow, the issi_dc working group defines a web page that supports searching according to Dublin Core element, associated with pictures, Nôm text and quốc ngữ transliteration. The style sheet also supports page turning, see *Fig. 5*, the Dublin Core page, and *Fig. 6*, content page for the book with an inventory id. ISSI HN 0443. These two figures show the same page format, and two ways of changing the page content:

a. **Change resource, searching by Dublin Core element**, for example, in *Fig. 5*, one can choose identifier inventory id ISSI HN 0443, to access the resource. Otherwise, the searching method is similar to the library searching method in use today, or googling, using the standard controlled vocabularies [15], or searching the entire content of the electronic resource, with morphological and synonym/antonym capacities.

b. **Change page** in a selected document, such as |← "first page", ← "previous page", → "next page", →| "last page", and *Turn to ___* (page), in *Figs. 5* and *6*. Here, we use a page format coded by PHP:

— displaying page image (e.g. page001.jpg in *Fig. 6*, and this picture can be isolated by clicking the mouse twice),

— catalogue information (file issi_HN_0443.php in *Fig. 6*),

— Nôm and Quốc ngữ transliteration information (e.g. page001.php in *Fig. 6*), etc.

— The scheme also works with commentaries and translation, using a standard file naming convention, thus allowing itself to be manipulated by the stylesheet, as demonstrated in the common page layout in *Fig. 5* (cover page) and *Fig. 6* (page 1). Thanks to the web stylesheet capability, the digital workflow tasks can be done independently, without geographical restrictions, and allowing outsourcing, when necessary.
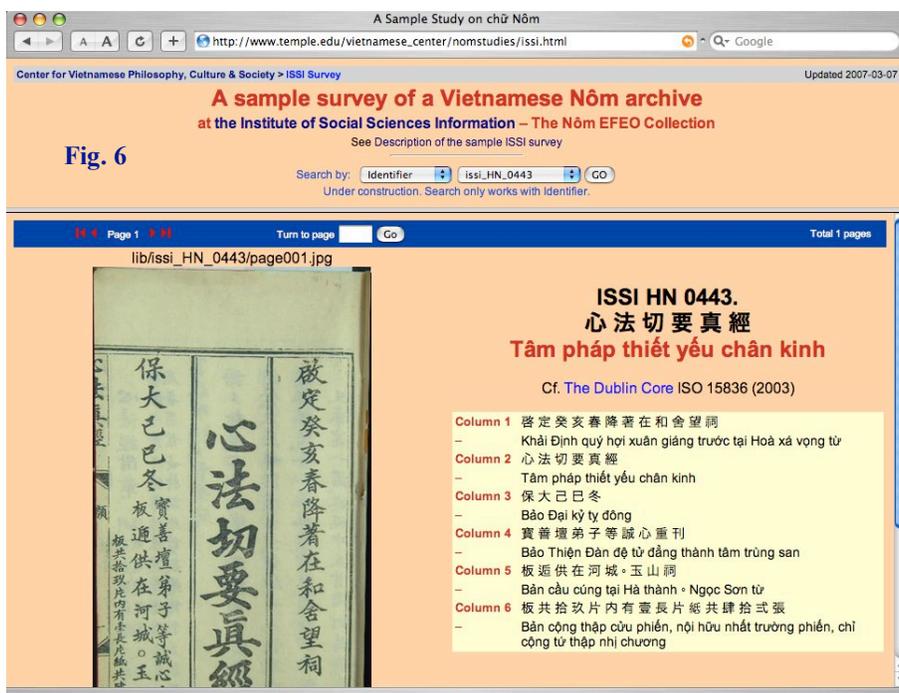
Fig. 5

Fig. 6

5. Conclusion

The Dublin Core issi_dc working group intends to continue this research on the following documents:

— ISSI HN 0443 Tâm pháp thiết yếu chân kinh (incomplete, only 1 page photographed)
— ISSI HN 0987 Book of Witchcraft (only 58 pages photographed)
— ISSI HN 1011 Record book of land application for cultivation (only 2 pages photographed)
— ISSI HN 1020 Pledges for the Temple for Đinh Ất Family Branch of Thái Bình Province, Kiến Xương District, Đông Nhuế Village (26 pages in total)
— ISSI TTTS 1261 Collection of Emperor Village Sainthood Decrees of Thái Bình Province, Kiến Xương Prefecture, Trực Định District, Thuy Lũng Canton (only 2 pages photographed)

These documents have been loaded onto the website for workflow activities according to *Fig. 3*, more imaging (image processing group), Dublin Core cataloguing, e.g. *Table 2* (Dublin Core group), Nôm

and quốc ngữ data entry, e.g. *Fig. 4* (Hán Nôm group) independently on the Temple website (IT group) at www.temple.edu/vietnamese_center/nomstudies/issi.html, under the direction of the managing group.

Reference

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* [Digital library], internal training slides: Set 1, General Issues, Set 2, Metadata standards MODS, METS and Dublin Core, Set 3: Extensible Store FEDORA and Open Archive OAI. Private communication.
7. Nôm Na Group. 2004. The Nôm Na Process: *Giúp đọc Nôm và Hán Việt* and Nôm on the web, *Proceedings of the International Nôm Conference*. Văn học Publishing House, Hà Nội. Pp. 396-410.
8. Nhóm Nôm Na. 2006. 6 phiên bản *Truyện Kiều*: những vấn đề văn bản học [6 versions of the *Tale of Kiều* and textual issues], presented at the *International Nôm Conference*, Huế, 6/2006.
9. Dũng, P.A. 2007. *Hệ thống phần mềm và trang web Hán Nôm HueCIT*, internal training slides of Thừa Thiên–Huế Center for Information Technology.
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. *Syllabeme and patterns of word formation in Vietnamese*, Ph.D. Thesis, New York University.

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* [Orthographic unit and characteristics of Vietnamese: the latin quốc ngữ script, the ideographic Nôm script, and Unicode/ISO IEC 10646], Vietnamese Unicode/ISO 10646 Committee [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.
— *3-digit Language Code,* ISO 639-2:1998, <www.loc.gov/standards/iso639-2/langhome.html>
— *Country Code,* ISO 3166, <www.din.de/gremien/nas/nabd/iso3166ma/>
— *Language Identification*, Internet RFC 3066. <www.ietf.org/rfc/rfc3066.txt>
— *Script Code*, ISO 15924. <www.unicode.org/iso15924/codelists.html>
— Internet Media Types. <www.isi.edu/in-notes/iana/assignments/media-types/media-types>
— *DCMI Type Vocabulary*. DCMI Proposal, 2000-07-11. <http://dublincore.org/documents/dcmi-type-vocabulary/>
— *Uniform Resource Identifiers*, URI: Generic Syntax, Internet RFC 2396. <www.ietf.org/rfc/rfc2396.txt>
— *Dublin Core Metadata for Resource Discovery*. Internet RFC 2413. <www.ietf.org/rfc/rfc2413.txt>
— *Getty Thesaurus of Geographic Names,* <www.getty.edu/research/tools/vocabulary/tgn>
— *Date and Time Format*, W3C Notes. <www.w3.org/TR/NOTE-datetime>.
16. 上海图书馆. 2006. 都柏林核心元数据元素集 1.1版: 参考描述, 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*.[Dublin Core: a web resource description standard]. Office of Newspapers & Magazines, Institute of Social Science Information. Personal communication.