

Global Issues in Vietnamese Nôm Preservation

Ngô Thanh Nhân

Center for Vietnamese Philosophy, Culture & Society

www.temple.edu/vietnamese_center/nomstudies

Global Temple Conference

Howard Gittis Student Center, Room 217D

Panel 4 – 10:10 AM November 13, 2007

Vietnamese basics

- Vietnamese is a tonal, monosyllabic language, belonging to the Mon–Khmer subgroup of the Austro–Asiatic language family.
- Each syllable has one of the 6 tones:

high	<i>lai</i> “mixed”	<i>lải</i> “tireless”	<i>lái</i> “to drive”
low	<i>lài</i> “jasmin”	<i>lãi</i> “interest”	<i>lạ</i> “return”

- Each syllable can be split into an initial consonant cluster *l-* and a rhyme *-ai*.

Two scripts in Vietnamese

- Vietnam has two national scripts: one is quốc ngữ, the other is喃.
- Quốc ngữ is a romanized script used in Vietnam since the 1920's.
- 喃 (reads nô) is an ideographic script used in Vietnam for over 1,000 years until the 1920's.
- They both represent Vietnamese language at different time in history.

Two scripts in Vietnamese

- A unit of utterance, and likewise, a phonetic unit of morphology, in Vietnamese is a syllable—called a **tiếng**.
- Each **tiếng** is represented by one **chữ**, an ideogram, in **nôm**.
- Each **tiếng** is represented by one **chữ**, a string of alphabet letters bounded by delimiters (blanks, punctuations, ...), in romanized **quốc ngữ**.
- Both scripts attempt to represent the sound of Vietnamese at different time in history.

Problem #1

- Hundreds of thousands of documents in Nôm, used in Vietnam for over 1,000 years, are now in danger of further destruction after more than 100 years of wars and neglect.
- Nôm documents and artifacts have been found languishing, unidentified, unprotected in many European and East Asian libraries, museums, private holdings, and all over Vietnam.

Problem #2

- The last national examination using Nôm was in 1919. The surviving Nôm scholars have nearly died out.
- The surviving Nôm scholars do not teach, because they did not go through modern pedagogical training.
- Thus, the wars, neglect, and the requirements of modern education jointly aggravate the situation.

Problem #3

- In universities, Nôm Studies belong to departments of literature, not Vietnamese Studies.
- Nôm teaching materials in Vietnam today are few. Official textbooks sorely lack important data about those ten centuries.
- The traditional teaching method lacks knowledge of pedagogy as well as educational materials.

Globalization of Nôm

- The preservation of the Vietnamese Nôm heritage is a desperate race against time.
- Globalization of Nôm studies seems to be an optimal way to preserve this heritage. It combines scientific rigor in research and education, and international human and material resources based on a multilingual computer context for the Nôm script.
- In an effort to address these problems, the Center for Vietnamese Philosophy, Culture & Society has established a Nôm Studies site, see www.temple.edu/vietnamese_center.

A monument in Nôm



November 13, 2007

Global Temple Conference

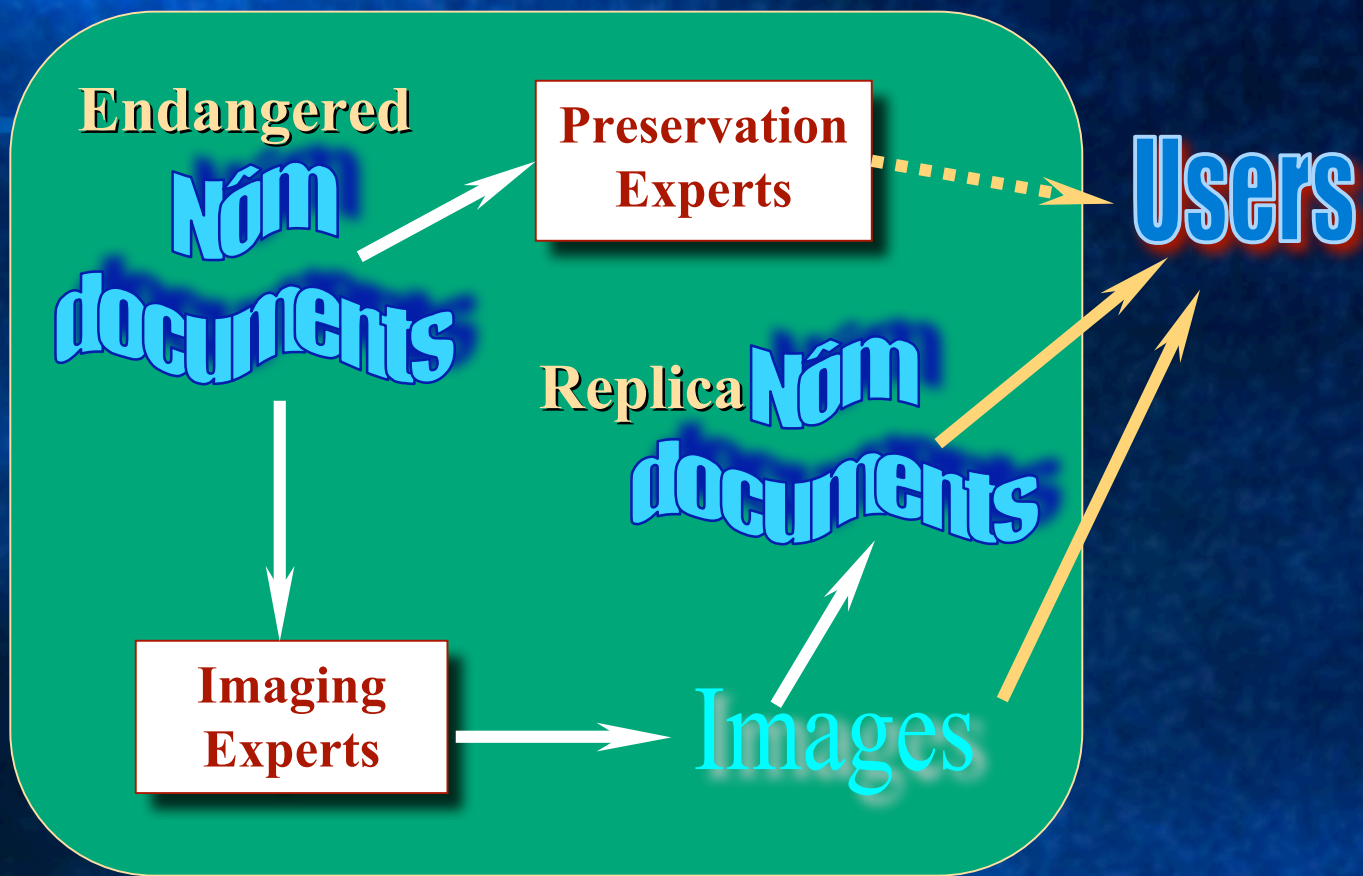
Preservation of Nôm

- Preservation of a heritage recorded in an endangered script usually implies physical preservation.



Preservation of Nôm

- Today, preservation also implies digitization, i.e. microfilm, digital photography or scanning.
- And, perhaps, a production of replicas accessible to users, while leaving the originals to preservationists.



Two types of digitization



打朱底魏遠
打朱底顛齧
打朱倣隻輪不返
打朱倣倣甲不還
打朱使知南國英雄
之有主

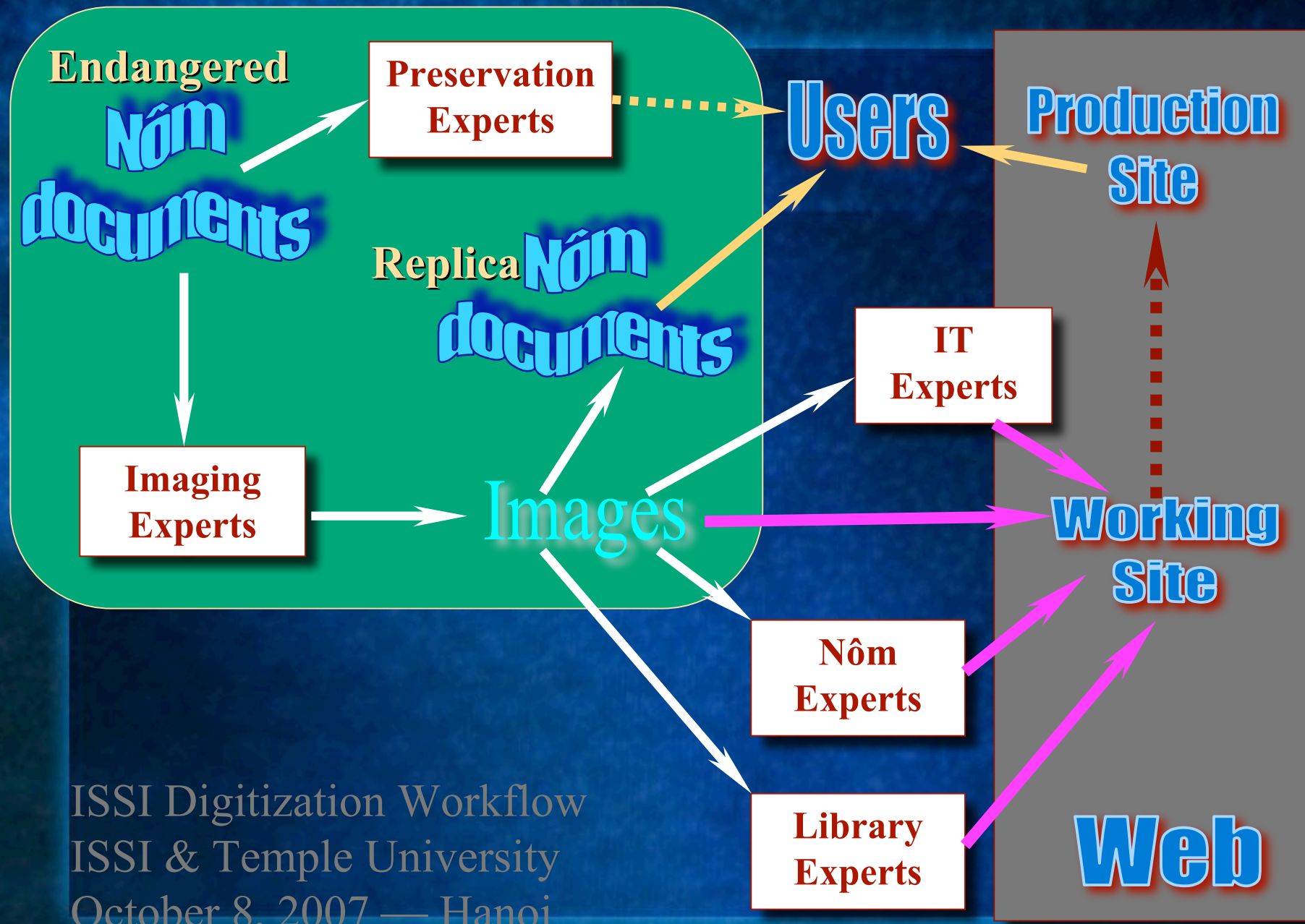
光忠阮惠

image

text

Preservation of Nôm

- Preservation by digitization can also mean bringing the ancient documents into the web platform, which can be searched and analyzed.
- This paper ventures a model, which allows preservation experts, imaging experts, Nôm experts, library information experts and IT experts to work together to achieve these goals.
- Our Center has organized two seminars this year at the Vietnam Institute of Social Science Information (ISSI) in Hanoi.



ISSI Digitization Workflow
 ISSI & Temple University
 October 8, 2007 — Hanoi

November 13, 2007

Global Temple Conference

Nôm script digitization

- Nôm script digitization implies building a Nôm ideogram repertoire, and ideogram and text knowledge bases (kBs).
- A language ideogram repertoire is a list of all unique ideograms seen in all written documents of a language.
- An ideogram knowledge base (ikB) is a list of ideograms, their shapes, their sounds, their descriptions, and their unique contexts in source documents.

A sample of the ikB

Unicode	Nôm	Quốc ngữ	“Radical”	URN
20CD2	啞	lời	口 khẩu	0030 7
53E3	口	khẩu	口 khẩu	0030 0
215F6	耂	trời	大 đại	0037 4
5929	天	thiên	大 đại	0037 1
4E0A	上	thượng	一 nhất	0001 2
5927	大	đại	大 đại	0037 0
4E00	一	nhất	一 nhất	0001 0
2E8A	卜	bốc	卜 bốc	0025 0

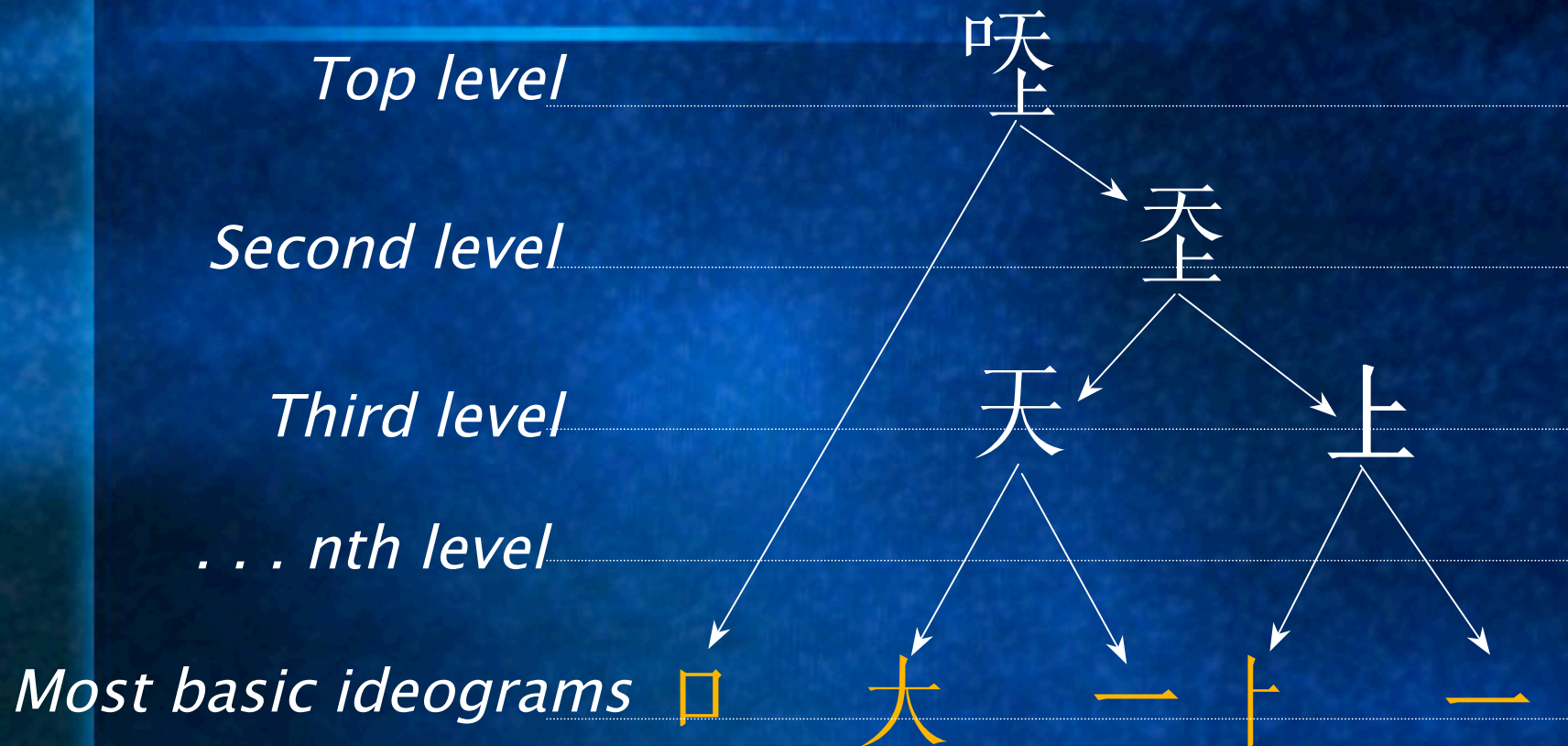
- Rows in color with 0 URN stroke count are “radicals”.
- URN: Unicode radical number.

Decomposition in the ikB

Unicode	Nôm	Quốc ngữ	Ideogram Description Pattern				
			UDC	E1	E2		
20CD2	啞	lời	𠄎	口	khẩu	𠄎	trời
53E3	口	khẩu					
215F6	𠄎	trời	𠄎	天	thiên	上	thượng
5929	天	thiên	𠄎	一	nhất	大	đại
4E0A	上	thượng	𠄎	卜	bốc	一	nhất
5927	大	đại					
4E00	一	nhất					
2E8A	卜	bốc					

- UDC: Unicode description character.
- Nôm in rows with empty UDC columns are **most basic ideograms**.

Ideogram description



Ideograms are formed by other ideograms.

OCR*

打朱底魏還
打朱底顛齧
打朱倣隻輪不返
打朱倣片甲不還
打朱使知南國英雄
之有主

打朱底魏還
打朱底顛齧
打朱倣隻輪不返
打朱倣片甲不還
打朱史知南國英雄
之有主
忠阮惠

Transliteration
into quốc ngữ

Đánh cho để dài tóc
Đánh cho để đen răng
Đánh cho nó chích luân bất phản
Đánh cho nó phiến giáp bất hoàn
Đánh cho sử tri Nam Quốc anh hùng
chi hữu chủ.

Quang Trung Nguyễn Huệ

Transliteration
into quốc ngữ

* OCR: optical character recognition, or image-to-text conversion.

OCR*

打朱底馘遠

打朱底顛齧

Đánh cho để dài tóc

Đánh cho để đen răng

Đánh cho nó chích luân b

Đánh cho nó phiến giáp b

Đánh cho sử tri Nam Quốc anh hùng

chi hữu chủ.

Quang Trung Nguyễn Huệ

打朱底馘遠

打朱底顛齧

打朱倂隻

打朱倂片甲不遠

打朱史知南國英雄

光中

有主

Transliteration
into quốc ngữ

Translation
into English

*We beat you because we like to wear our hair long.
 Beat you because we like to blacken our teeth.
 Beat you, so none of your war chariots could run off.
 Beat you to keep your weapons from going home.
 Beat you so history knows the South has its own king.
 Nguyễn Huệ, Emperor Quang Trung, 1789
 (translated by John Balaban)*

Nôm digitization study

- The ISSI digitization model is being tested in the context of Vietnam. It is a model for web surveys of the Nôm archive at the Vietnam Institute for Social Science Information in a joint project with the Center for Vietnamese Philosophy, Culture & Society here at Temple.

Nôm studies & Dublin Core

- The ISSI–Temple/Center joint project has established an “ISSI DC” group—a study group on digitization of the ISSI Nôm archive using the Dublin Core, a World Wide Web library information standard.
- The ISSI DC group consists of:
 - (1) a Nôm ideographic team,
 - (2) a Dublin Core library info. team,
 - (3) a digital image team, and
 - (4) an IT web design team.
- The Temple/Center hosts the ISSI DC web page.

Nôm studies at the Center

- The Temple/Center focuses on Vietnamese Studies, of which Nôm Studies is a part.
- Nôm studies includes multi-disciplinary research on Vietnamese historical records written in Nôm, i.e. for over 1,000 years before the 1920's.
- The Temple/Center hosts a Nôm Studies web page to publish academic research on Nôm.
- The Temple/Center plans a mini Nôm conference next April 2008 and a Nôm course for beginners.

References

- Antelman, K., Lynema, E. & Pace, A.K. Toward a twenty-first century library catalog, *Information Technology & Libraries* (2006): 128–139.
- Eberthart, George M. ed. 2006. *The whole library handbook*. Chicago: American Library Association.
- Martin, Lowell A. 1996. *Organizational structure of libraries*. London: The Scarecrow Press.
- Schutze, Gertrude. 1972. *Information and library science source book*. Metuchen, NJ: The Scarecrow Press, Inc.

For further information, contact Dr. Ngo Thanh Nhan, Visiting Research Scholar, Center for Vietnamese Philosophy, Culture, and Society at nhan@temple.edu.