

Do.M.En.I.

Dongba Manuscripts Encoding Initiative

Introduction to the script and resume

UNESCO recognized to Lijiang scenery and city, Naxi people and culture, Dongba tradition and pictograph manuscripts literature as part of Human World Heritage.

According to UNESCO pictograph writing system and manuscripts pictograph art is one of peculiar CULTURAL EXPRESSION and ARTISTIC MANIFESTATION OF NAXI PEOPLE.

Pictograph writing system and Dongba literature are then Human World Heritage too, thus World Wide Web technology appears as one of the powerful instrument to render such human kind treasure enjoyable and available in whole world.

Diffusion of Dongba pictograph manuscript tradition is acculturation to Naxi people cultural richness and China ethnic variety.

Do.M.En.I. is a research Project focused about:

- **defining and building a codify schema for encoding Dongba pictograph manuscripts**
- **implementing a World Wide Corpus of encoded Dongba Manuscripts**

Do.M.En.I. - Dongba Manuscript Encoding Initiative project originates and takes form by Dongba pictograph manuscript tradition and digital humanities study, and is focused about implementing an on-line encoded corpus of Dongba manuscripts as a World Wide enjoyable resource for Dongba manuscript and Naxi pictograph literature.

Do.M.En.I. project target is thus implementation of a growing and up-to-date open on-line archive of Dongba texts as annotated corpus, following L.C. and W3C standards recommendations.

Do.M.En.I. has to struggle against some difficult factors for its implementation:

- Peculiarity of Dongba manuscript, wrote in pictograph and Geba syllable-graphic writing systems
- Few just-implemented and available digital reproduction of Dongba manuscripts
- Heterogeneity of such few available digital resources
- Objective difficult in retrieving original manuscripts for making new digital resources
- Dis-information of available manuscript in Europe and Western world

I feel just listed problems as incentive in hard-working for Do.M.En.I. implementation and realization.

Why Do.M.En.I. ?

World Wide web technology at today should be seen as the wider library to be consulted for information retrieval and documents browsing, especially in Academic world for exchanging of material and up-to-date information which run faster then every paper-publication or review.

Many digital humanities application is dedicated to texts browsing and documents retrieving, and cover large range of kind of information: from daily newspapers to linguistic annotated corpora or more specialized archives.

Plenty of digital texts and their extreme heterogeneity pushed W3C consortium (among other world wide question) to define some base guidelines for digital encoding of text and corpora digital encoding. at today W3C recommends TEI guidelines for digital texts encoding and CES for corpora encoding.

Do.M.En.I., project of diffusion and acculturation about Dongba pictograph manuscript tradition and Naxi people culture, follows and accords to W3C standards recommendations because it's the best way to open such cultural and artistic treasures to academic world and all members of human kind.

Do.M.En.I. thus operates in retrieving of just-available on-line resources re-organizing them in a CES according corpus, proceeding to single manuscripts encoding according to TEI

recommendations.

At today two are the only available on-line resources for Naxi pictograph manuscripts tradition, both displaying some tens of digital reproduction of manuscripts pages:

- **Library of Congress, a selections from the Naxi Manuscript Collection:**
<http://memory.loc.gov/intldl/naxihtml/naxihome.html>
→ 24/06/2009, with stored in, available as images.
- **Naxi manuscripts archives of Harvard University Library O.A.S.I.S. project** (*on-line Archival Search Information System*):
http://oasis.lib.harvard.edu/oasis/deliver/deepLink?_collection=oasis&uniqueId=hyl00002
→ 24/06/2009, with manuscripts stored in, available as images

they have heterogeneous standards for digital reproduction of manuscripts, moreover both corpora is just stops to "0 level" of text encoding. (Lenci et al., 2005: 57) , for Dongba pictograph text reproduction of manuscripts pages, sequences of images without any deeper structural, meta-structural, linguistic, syntactic and semantic annotation.

Do.M.En.I. projects if focused in building a complete (and open for up-to-dating initiatives) resources as annotated corpus of Dongba manuscripts, and this plays for:

- **safeguard and conservation of Dongba manuscripts and Naxi pictograph literature**
 - conservation of Dongba manuscripts as non-perishable documents
 - availability of virtual facsimiles for academic study
- **convergence and unification of just-existing on-line resources**
 - preventing dispersion of resources (cfr below: **Dongba manuscripts, Naxi pictograph tradition literature corpus**)
 - preventing heterogeneity of material (*idem*)
 - linking of different institution for more easy cooperation (*idem*)
- **implementation of new resources according to W3C standards**
 - Dongba manuscripts census in world libraries (cfr below: **Dongba manuscripts, Naxi pictograph tradition literature corpus**)
 - creation of an unique index meta-catalog to supply at Dongba manuscripts spread out (*idem*)
 - incentive of digital encoding initiative of representative sample for just-paper Dongba manuscripts collections (*idem*)
 - programming of initiatives aimed to compensate Dongba culture and Naxi literature at today unrepresented (*idem*)
- **diffusing and enjoying Dongba tradition and Naxi pictograph literature**
 - acculturation to Dongba tradition
 - acculturation to Naxi culture
 - acculturation to China ethnic variety
- **study of Dongba pictograph writing system**
 - as the only alive pictograph writing system
 - in relation to others (death) pictograph-hieroglyphics writing system
 - in relation to modern iconic media
- **study of Naxi language**
 - definition, according to Naxi people authorities, of a standard pinyin system for latinization of Naxi and transliteration of Dongba pictographs and Geba syllable-graphs
 - creation of equivalence plates about latinization for previous mains work about Naxi, Dongba and Geba
 - implementation of a Naxi → Dongba lexicon (ancient Naxi) with linguistic annotations
 - implementation of a Naxi modern language lexicon with linguistic annotation
 - implementation of cross ancient Naxi <-> today's Naxi language study for pattern identification

How to Do.M.En.I. ?

Do.M.En.I. project work should be resumed in two focal targets:

1. defining and building a codify schema for encoding Dongba pictograph manuscripts according to TEI
2. implementing a World Wide Corpus of Dongba Manuscripts according to CES

For TEI and CES standard recommendations please cfr official guidelines respectively

- TEI – Text Encoding Initiative: <http://www.tei-c.org>
- CES – Corpus Encoding Standard: <http://www.cs.vassar.edu/CES/>

Dongba manuscript encoding

Dongba manuscripts encoding in few words consists in projecting and implementing an useful encoding schema calibrated for Dongba manuscripts and according to TEI standards.

Among many excellent samples of calibrated encoding schema TEI accorded here we're quoting ME.NO.T.A. - Medieval Nordic Text Archive <http://www.menota.org> initiative, “*a network of leading Nordic archives, libraries and research departments working with medieval texts and manuscript facsimiles. The aim of Menota is to preserve and publish medieval texts in digital form and to adapt and develop encoding standards necessary for this work. The archive will contain texts in the Nordic languages as well as in Latin.*” (http://www.menota.org/index_en.page → 2009, June 27th)

Making treasure of ME.NO.T.A. experimenting some drafts solution on www.xiulong.it for

- Leabhar Cheanannais The Book of Kells - Evangelario di San Colombano, in Italian: <http://www.xiulong.it/Kells> → 2009, June 27th
- TEI standard encoding project of one Dongba manuscript, in Italian <http://www.xiulong.it/Dongba/main/index.xml> → 2009, June 27th

both are attempts, firming up of drafts focused about ancient manuscripts study approaching by human technologies; Do.M.En.I. project springs out also from this drafts and from them are enriched.

Do.M.En.I. works in will consist in building up an ENCODING SCHEMA for Dongba manuscripts encoding ad Corpus making, and such work consists in:

- **detecting of characteristics to be marked-up in Dongba manuscripts**
 - what kind of mark-up
 - meta-structural
 - structural
 - linguistic
 - syntactic
 - semantic
 - others information to be marked-up
- **implementation of Do.M.En.I. markers to explicit what previous detected**
 - adopting TEI markers which are just good for Do.M.En.I. schema
 - Xml implementation of new markers to supply and integrate TEI's eventually lacks working over Dongba manuscripts

Once decided what needs to be marked-up and markers is possible to:

- **implementing an *domeni.dtd*, alias Xml Document Type Definition for Do.M.En.I. encoding initiative,**
 - available as universal DTD for encoding whatever Dongba manuscript
 - compatible with TEI DTDs
- **implementation of an Xml markup language for Dongba manuscript encoding**
 - implementation of Xml markers repertory (elements, attributes and identities) well-formed and validated to *domeni.dtd*
 - downloadable empty model ready-to-use (an Xml empty file but validate and well

- formed to *domeni.dtd*) for starting encoding Dongba manuscript
- Do.M.En.I. guideline with markers repertory recommendations

Implementation of Dongba manuscripts encoding schema is not limited or just related to creation of a TEXTS ENCODING INITIATIVE, but assuming a fundamental role of

- Pictograph writing system
- presence of illuminated portion of manuscript

it evident that encoding Dongba Manuscript Corpus consists also in standardized digitals facsimile of manuscripts that has to be implemented with TEI recommendations for facsimiles and manuscript encoding.

Encoding manuscripts also has to explicit

- OBJECTIVE DIFFERENCES BETWEEN WHAT IS WROTE IN PICTOGRAPHS on MANUSCRIPT and
- WHAT IS EFFECTIVE SPELLED BY DONGBAS PRIESTS WHILE "chanting" MANUSCRIPT during ritual celebration

because in manuscripts, oral part of speech and wrote pictographs aren't equivalent.

Dongba manuscripts, Naxi pictograph tradition literature corpus

One of the most distinctive features of Naxi culture is Dongba tradition and related pictographic texts wrote in form of some thousands of manuscripts: according to Jackson and Anshi (Oppitz M., Hsu E., 1998: 240) 5.118 manuscripts actually held in libraries of Europe and America, as:

- British Library
- India Office Library in London
- John Rylands Library in Manchester
- Staatsbibliothek in Berlin
- Library of Congress in Washington D. C.
- Chinese-Japanese Library of Harvard-Yenching Institute in Boston, Massachusetts

Manuscripts could be divided in three main kinds:

1. **books for chanting religious ceremony**, containing religious texts, tales, formulas and mantras to be chanted during rituals performing
2. **index books**, alias guides to how perform rituals and/or ceremonies, fingering the paraphernalia to be used and listing the manuscripts to be chanted
3. **books of divination**, determining in which circumstances which ceremony was to be performed.

Jackson and Anshi study (Oppitz M., Hsu E., 1998: 237 - 274) focused about style of manuscripts writing and schools of Naxi Dongba authors pointed out 5 main styles of writing manuscripts labeled as A,B,C,D and E:

- **A group:** manuscripts are related to *Baisha's* school (see below), of profitable production center of fine artistic styled manuscripts. Jackson and Anshi divided A style in 5 sub-category:
 - A1: authors, who coincided with Rock's *Dto-la* Dongba brothers, possessed very artistic skills, superior to common Dongba writer, with colored miniatures onto 1st page of manuscripts. Most of the cover-titles were set in decorated frame on the top by draperies with a precious object in the middle as:
 - pearl in flames,
 - eight-spoked wheel,
 - conch-shell
 - purification *Amirta* vase

- love-knot of good-luck knot.

Iconographically associable also for incipits as:

- incipit pictograph of a Dongba priest miniature depicted three-lobed crowned while sitting and chanting.
- incipit symbol at the beginning of manuscript as the beginning of Tibetan books, representing the Tibetan letter ཨ [á] which refers to the great god in the Naxi pantheon.

- A2: wrote by Dongba doted of very skilled artistic features: Jackson and Anshi states about *Ho Ho-shou* as the author, Dongba from the village of Nankow (Nankou Xin Cun, Naxi *La 'ts 'u wua*) south-east of Lijiang
- A3, coinciding with Rock's *He Huating* Dongba, who taught Rock to read religious texts, also the drawer of pictographs in Rock's dictionary and other publications. He also wrote some title-pages and added them to some ritual manuscripts written by different hand.
- A4 fine and neat drawings unique in character
- A5 similar with A4, with neat drawings, but coinciding with Rock's *Dto dsu* Dongba.

- **B group:** presents more complicated pictures, and authors used more elaborated symbols for decorating title-pages; decorative draperies differs iconographically from A type.

According to Rock, manuscripts collected in B group were produced from *Mun-shua-wua* village, identified by Jackson and Anshi with *Runanhua* village of *Tai 'an* township: they also divides B group in 9 sub-groups because of their iconography variety of decorations.

- **C group:** C type presents titles very more elaborated than A and B, being decorated with six draperies, and among them Jackson and Anshi identified 7 different writers hands.
- **D group:** they aren't not sophisticated as those of A, B and C groups, with title-pages set side-way. Many of this manuscripts are divination books and mostly came from *Baidi* and *Baoshan*.
- **E group:** presents non-elaborated title ornamentation, made by authors not as skilled as A, B, and C authors: Jackson and Anshi states they probably were novices.

Jackson and Anshi also studied style groups distributions and identified 4 main regional nucleus of manuscripts production, and they called them as "*school*", alias focal centers of distribution of manuscripts:

- **Baisha:** including Dongba of Lijiang plain, at today economic and cultural center of Naxi community. Baisha school appears to be specialized in production of book for chanting religious and ceremony, and they used to write both in pictographs and syllable-graphs. According to Jackson & Anshi Baisha center comprises 5 main areas in 2 or 3 miles around Lijiang:
 - *Baisha* market place, north of *Lijiang*
 - *Wenbi* and *Changshui* villages in *Lashi* township, south-west of *Lijiang*
 - *Guifeng* and *Liangmei* township, east of *Lijiang*
 - *Wutai* township, south of *Lijiang*
 - *Lijiang* township itself, also named *Dayanzhen*

- **Tai'an and Ludian**

Tai'an township is south-west of *Lijiang*, especially known for Dongba activities in *Runanhua* village. According to Dongba tradition, the founder of Dongba religion *Shilo* stayed for sometimes in a cave called “spiritual cave of *Shilo*” just in front of *Runanhua* village.

Such school included also:

- Dongba from *Ludian* town itself, west of *Lijiang*, near *Weixi* country
- Dongba from *Yilong* village and *Tacheng* township, near *Ludian*
- Dongbas of *Tai'an* and *Ludian* wrote both in pictographic and syllable-graphic writing systems.

- **Baidi**

located in *Sanba* township of *Zhongdian* country, north of *Lijiang*, beyond Gold Sand River. According to Naxi tradition, *Shilo* had his seat in *Baidi* itself: there's a spiritual cave here dedicated to, and *Baidi* is considered to be the place of origin of Naxi/Dongba religion, infact it was a destination for a Dongba adepts to went here in pilgrimage in order to become Dongba masters.

Baidi's manuscripts corpus didn't include texts dedicated to the great ceremony *Har-la-llu-k'o* for suicides as in their rituals, *Baidi's* Dongba didn't perform sword dancing as they not use syllable-graphics.

- **Baoshan**

school comprising townships of

- *Fengke*
- *Mingyin*
- *Daju*
- *Baoshan* itself

all in the north of *Lijiang*, within the loop of the Gold Sand River.

Divination was very popular among *Baoshan's* Dongbas, and they used to write – as *Baidi's* Dongbas – only by pictographs.

Corpus Encoding

Dongba pictograph manuscripts tradition number at today several thousands of manuscripts wrote on traditional *Lijiang* papyrus by pictograph and Geba writing system.

Such documents constitute the whole literature of Naxi – Dongba cultural tradition, and they constitute a textual corpus, which could be organized by major and minor ceremonies, assuming them as index of manuscripts corpus.

Such corpus should be considered as ecologic setting where pictograph writing system live still today a setting characterized by an evident peculiarity: Dongba manuscripts are all about Dongba religion and Naxi tradition, thus Dongba Corpus is extremely SPECIALISTIC, or VERTICAL CORPUS, because it's made of texts belonging to both:

- A SPECIFY LINGUISTIC VARIETY
- A PECULIAR THEMATIC DOMAIN

In such vertical corpus previously resume of Dongba manuscripts subdivision and classification

1. kind of manuscripts
2. style
3. regional distribution

represents basilar concept which Do.M.En.I. has to reflect and represents in its corpus encoding.

Other focal point, as introduced upper, Do.M.En.I. is conceived according to C.E.S. - Corpus Encoding Standard, which, in turn, is compatible with TEI standard.

In 1st kind of manuscripts, alias **books for chanting religious ceremony**, is then necessary

to specify which ceremony or ceremonies manuscripts are dedicated to, and this characteristic have to be reflected in Do.M.En.I. implementation.

Corpus Encoding thus should reflect such organization, and has to be structured in a hierarchical succession of nodes and leaves, for instance:

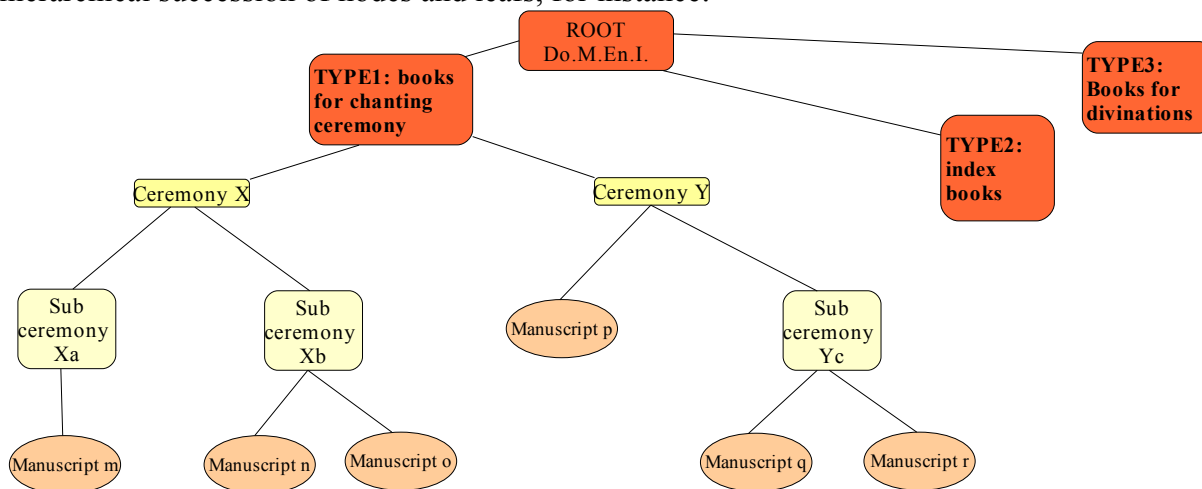


Diagram could well be reproduced in Xml well-formed tree:

```

<domeni> // root node
  <type id='1' kind='chant_cer'> // begin of 1st kind of manuscript, alias book for chanting ceremonies
    <ceremony type='major' name='X'> // beginning of X major ceremony's manuscripts encoding
      <ceremony type='minor' name='Xa'> // beginning of Xa sub-ceremony's manuscripts encoding
        <manuscript id='m'> // beginning of "m" manuscript encoding
          ...
        </manuscript>
      </ceremony>
      <ceremony type='minor' name='Xb'> // beginning of Xb sub-ceremony's manuscripts encoding
        <manuscript id='n'> // and Xb manuscripts encoding...
          ...
        </manuscript>
        <manuscript id='o'>
          ...
        </manuscript>
      </ceremony> // end of Xb sub-ceremony
    </ceremony> // end of X ceremony
    <ceremony type='major' name='Y'> // beginning of Y major ceremony
      <manuscript id='p'> // beginning of a "p" manuscript related to Y major ceremony without being
        ... // part of any sub-ceremony. P manuscript could easy be an index
        </manuscript> // or divination manuscript
      <ceremony type='minor' name='Yc'> // beginning of Yc minor ceremony
        <manuscript id='q'> // and Yc's manuscripts "r" and "q"
          ...
        </manuscript>
        <manuscript id='r'>
          ...
        </manuscript>
      </ceremony> // end of Yc minor ceremony
    </ceremony> // end of Y major ceremony
  </type> // end of 1st type of manuscripts encoding
  <type id='2' kind='index'> // begin of 2nd kind of manuscript, alias index book
    ...
  </type> // end of 2nd kind of manuscripts encoding
  <type id='3' kind='divination'> // begin of 3rd kind of manuscript, alias divination books
    ...
  </type> // end of 3rd kind of manuscript encoding
</domeni> // end of Do.M.En.I. corpus

```

Another level of mark-up apply-able to Do.M.En.I. corpus should describes bigger groups of ceremonies pooled by some common meta-features as:

- all ceremonies belonging to worship to heaven,
- all ceremonies belonging to worship to earth,
- all ceremonies belonging to worship of ancestors,

- all ceremony dedicated to demon suppressing
etc...

Moreover, another kind of featured mark-up of Dongba literature that Do.M.En.I. has to implements is linguistic marking-up of Dongba corpus, distinguishing at first among 3 basic cases:

1. just pictograph manuscripts,
2. just Geba syllable-graphic manuscripts,
3. both writing systems manuscripts (maybe the wider set)

Implementation of Do.M.En.I. corpus markers to explicit previous characteristics has to

- adopts CES markers which are just good for Do.M.En.I. schema
- implements Xml new markers to supply and integrate CES's eventually lacks working over Dongba manuscripts corpus

and then proceeding with identification and implementation of markers for encoding a Corpus able to contains Dongba manuscripts by

- implementation of a DTD for Dongba Corpus building
- implementation of DTD's related markup language for Dongba Corpus Encoding



Xiulong

Stefano Zamblera
2009, July 27th