Some Remarks Concerning Digital Editions of Premodern Chinese Texts

Christian Wittern

2010-03-26

Overview of the presentation

- Some general remarks about digital text

- Some general remarks about digital text
- Description of the project and its workflow

Overview of the presentation

- Some general remarks about digital text

- Description of the project and its workflow

- Plans for enhancement towards a framework for working with digital text that satisfy rigid requirements for scholarly use

Digital text

- Digital texts are still at the incunable stage of development

Digital text

- Digital texts are still at the incunable stage of development
- so far we have barely scratched the surface of their possibilities

Digital text

- Digital texts are still at the incunable stage of development
- so far we have barely scratched the surface of their possibilities
- there is no "given" architecture, infrastructure or publication model

Digital text

- Digital texts are still at the incunable stage of development
- so far we have barely scratched the surface of their possibilities
- there is no "given" architecture, infrastructure or publication model
- our imagination and vision brings it into existence

Digital text

- Digital texts are still at the incunable stage of development
- so far we have barely scratched the surface of their possibilities
- there is no "given" architecture, infrastructure or publication model
- our imagination and vision brings it into existence
- they are formed according to our wishes, but

Digital text

- Digital texts are still at the incunable stage of development
- so far we have barely scratched the surface of their possibilities
- there is no "given" architecture, infrastructure or publication model
- our imagination and vision brings it into existence
- they are formed according to our wishes, but
- Web publication is **not** *the* answer

Digital text

- Digital texts are still at the incunable stage of development
- so far we have barely scratched the surface of their possibilities
- there is no "given" architecture, infrastructure or publication model
- our imagination and vision brings it into existence
- they are formed according to our wishes, but
- Web publication is **not** *the* answer
    - it leads to disengagement of the text and its readers

Three phases in the life of a digital text

- (1) creation

Three phases in the life of a digital text

- (1) creation
    - birth

Three phases in the life of a digital text

- (1) creation
    - birth
    - => establishment of base text and its variants, its characters and basic compositional units

Three phases in the life of a digital text

- (1) creation
    - birth
    - $=>$ establishment of base text and its variants, its characters and basic compositional units
- (2) enhancement

Three phases in the life of a digital text

- (1) creation
    - birth
    - => establishment of base text and its variants, its characters and basic compositional units
- (2) enhancement
    - education

Three phases in the life of a digital text

- (1) creation
  - birth
  - => establishment of base text and its variants, its characters and basic compositional units
- (2) enhancement
  - education
  - => identification and collation of other features of the text, annotation and commentary on the text

Three phases in the life of a digital text

- (1) creation
    - birth
    - => establishment of base text and its variants, its characters and basic compositional units

- (2) enhancement
    - education
    - => identification and collation of other features of the text, annotation and commentary on the text

- (3) sharing (publication)

Three phases in the life of a digital text

- (1) creation
    - birth
    - => establishment of base text and its variants, its characters and basic compositional units

- (2) enhancement
    - education
    - => identification and collation of other features of the text, annotation and commentary on the text

- (3) sharing (publication)
    - work (making oneself useful in society)

Three phases in the life of a digital text

- (1) creation
    - birth
    - => establishment of base text and its variants, its characters and basic compositional units

- (2) enhancement
    - education
    - => identification and collation of other features of the text, annotation and commentary on the text

- (3) sharing (publication)
    - work (making oneself useful in society)
    - => distribution model, rights, access

Three phases in the life of a digital text

- (1) creation
  - birth
  - => establishment of base text and its variants, its characters and basic compositional units
- (2) enhancement
  - education
  - => identification and collation of other features of the text, annotation and commentary on the text
- (3) sharing (publication)
  - work (making oneself useful in society)
  - => distribution model, rights, access
- in reality, (2) and (3) form a cycle that is repeated indefinitely

- The most important Daoist textual collection is the Daozang 道藏 (DZ)

The Daozang 道藏 ...

- The most important Daoist textual collection is the Daozang 道藏 (DZ)

- The DZ has a complex and long history, starting from compilations in ca 400 AD

- The most important Daoist textual collection is the Daozang 道藏 (DZ)

- The DZ has a complex and long history, starting from compilations in ca 400 AD

- The oldest completely preserved edition was printed in 1445 (during the Zhengtong period, thus Zhentong Daozang)

The Daozang 道藏 ...

- The most important Daoist textual collection is the Daozang 道藏 (DZ)

- The DZ has a complex and long history, starting from compilations in ca 400 AD

- The oldest completely preserved edition was printed in 1445 (during the Zhengtong period, thus Zhentong Daozang)

- The DZ was compiled on imperial orders under the leadership of Daoist clerics

The Daozang 道藏 …

- The most important Daoist textual collection is the Daozang 道藏 (DZ)

- The DZ has a complex and long history, starting from compilations in ca 400 AD

- The oldest completely preserved edition was printed in 1445 (during the Zhengtong period, thus Zhentong Daozang)

- The DZ was compiled on imperial orders under the leadership of Daoist clerics

- it contains about 1500 texts with more ca. 5300 scrolls (26 Mio characters)

- A digest was created in the Qing period, initiated probably by Peng Dingqiu 彭定求 (1645-1719), but mainly compiled by Jiang Yupu 蒋予蒲 (1755-1819)

… and the Daozang jiyao 道藏輯要 (DZJY)

- A digest was created in the Qing period, initiated probably by Peng Dingqiu 彭定求 (1645-1719), but mainly compiled by Jiang Yupu 蒋予蒲 (1755-1819)
- The DZJY was compiled by a group of lay people interested in Daoism

… and the Daozang jiyao 道藏輯要 (DZJY)

- A digest was created in the Qing period, initiated probably by Peng Dingqiu 彭定求 (1645-1719), but mainly compiled by Jiang Yupu 蔣予蒲 (1755-1819)
- The DZJY was compiled by a group of lay people interested in Daoism
- A revised reprint was prepared in Sichuan in the early 20th century

… and the Daozang jiyao 道藏輯要 (DZJY)

- A digest was created in the Qing period, initiated probably by Peng Dingqiu 彭定求 (1645-1719), but mainly compiled by Jiang Yupu 蔣予蒲 (1755-1819)
- The DZJY was compiled by a group of lay people interested in Daoism
- A revised reprint was prepared in Sichuan in the early 20th century
- About 60% of the material is from DZ, 40% has been added newly

… and the Daozang jiyao 道藏輯要 (DZJY)

- A digest was created in the Qing period, initiated probably by Peng Dingqiu 彭定求 (1645-1719), but mainly compiled by Jiang Yupu 蔣予蒲 (1755-1819)
- The DZJY was compiled by a group of lay people interested in Daoism
- A revised reprint was prepared in Sichuan in the early 20th century
- About 60% of the material is from DZ, 40% has been added newly
- The differences between the DZ and DZJY, the newly added material and the two main strata of surviving editions require careful study

… and the Daozang jiyao 道藏輯要 (DZJY)

- A digest was created in the Qing period, initiated probably by Peng Dingqiu 彭定求 (1645-1719), but mainly compiled by Jiang Yupu 蒋予蒲 (1755-1819)
- The DZJY was compiled by a group of lay people interested in Daoism
- A revised reprint was prepared in Sichuan in the early 20th century
- About 60% of the material is from DZ, 40% has been added newly
- The differences between the DZ and DZJY, the newly added material and the two main strata of surviving editions require careful study
- A project was conceived to support this study, genereously supported by CCK and JSPS

- Establish and publish a new critical edition of the DZJY

- Establish and publish a new critical edition of the DZJY
  - including a digital one in many forms

Aims of the DZJY project

- Establish and publish a new critical edition of the DZJY
    - including a digital one in many forms
    - lay the foundation for further study of the collection

- Establish and publish a new critical edition of the DZJY
  - including a digital one in many forms
  - lay the foundation for further study of the collection

- Compile a critical, comparative catalog of the DZJY

Aims of the DZJY project

- Establish and publish a new critical edition of the DZJY
    - including a digital one in many forms
    - lay the foundation for further study of the collection

- Compile a critical, comparative catalog of the DZJY

- Publish a *Companion to the Daozang jiyao* similar to the 3 Vols. *Historical Companion to the Daozang* by Schipper/Verellen

Aims of the DZJY project

- Establish and publish a new critical edition of the DZJY
    - including a digital one in many forms
    - lay the foundation for further study of the collection

- Compile a critical, comparative catalog of the DZJY

- Publish a *Companion to the Daozang jiyao* similar to the 3 Vols. *Historical Companion to the Daozang* by Schipper/Verellen

- The following will discuss the progress and methodology of the digital DZJY

叢生姓　牲生㸦　㸦生产　牲生也　比相生也

㸦生菲　菻生烎　烎生器　菻对烎　烎对器　器对㸦

尸对卤　卤对卉　卉对卉　卉对婶　零对婶

册对囷　困对蠱　蠱对茷　茷对姓　姓对㸦

輪範屬先天之學

齐牲一道姓卉者言地可以勝天而人更可勝地盇地近而

天遠而人更近於天地也他则卉為凶

齐牛陽道生天生五行生星斗生金玉生龐馬生日月生男女

亓卉難分為二而生養萬物培植一氣則一也

牲生陰道生地生山川生草木生金玉生飛禽生雲霧生走獸

Yupu and Chongkan edition face to face

Right panel

傳一言許肯更無難數篇奇怪交人手一夜挑燈讀不了曉來
日早繞看畢不覺自醉如恍惚惚之中見有物狀如日輪明
突兀自言便是丹砂精宜向鼎中烹凡質本來不化真化
真須得真中物不用鉛不用汞還丹須向爐中種玄中之玄號
一粒名千般中有一為丹母火莫然水莫凍修之鍊之須珍
重直待虎嘯折巔峯驪龍奪得玄珠弄龍吞立實忽昇飛飛龍
被我捉來騎一鳶上朝歸碧落碧落廣闊無東西無曉無夜無
日月無寒無暑無四時自從修到無為地始覺奇之又怪之

贈劉方處士

呂帝詩集

室集六

Left panel

傳一言許肯更無難數篇奇怪文人手一夜挑燈讀不了曉來
日早繞看畢不覺自醉如恍惚惚之中見有物狀如日輪明
突兀自言便是丹砂精宜向鼎中烹凡質本來不化真化
真須得真中物不用鉛不用汞還丹須向爐中種玄中之玄號
一粒名千般中有一為丹母火莫然水莫凍修之鍊之須珍
重直待虎嘯折巔峯驪龍奪得玄珠弄龍吞立實忽昇飛飛龍
被我捉來騎一鳶上朝歸碧落碧落廣闊無東西無曉無夜無
日月無寒無暑無四時自從修到無為地始覺奇之又怪之

贈劉方處士

呂帝詩集

- Modern character encodings for Chinese are developed with the modern understanding of characters

Digitizing of Chinese text

- Modern character encodings for Chinese are developed with the modern understanding of characters
- This makes it very difficult to use for premodern text

Digitizing of Chinese text

- Modern character encodings for Chinese are developed with the modern understanding of characters

- This makes it very difficult to use for premodern text

- It is thus next to impossible to study changes in orthography and character usage in digital text

Digitizing of Chinese text

- Modern character encodings for Chinese are developed with the modern understanding of characters

- This makes it very difficult to use for premodern text

- It is thus next to impossible to study changes in orthography and character usage in digital text

- A new model for digitizing Chinese text is needed

treasure, precious; valuable; p5357

small thing, child; young animal; p5338

a walled village; protect, safeguard; p5345

swaddling cloth; infancy; p5420

Establish a new model for digital Chinese text

- Facsimile **or** Transcription -> Facsimile **and** Transcription

Establish a new model for digital Chinese text

- Facsimile **or** Transcription -> Facsimile **and** Transcription
- Not character codes, but position of the characters in the text as primary reference

Establish a new model for digital Chinese text

- Facsimile **or** Transcription -> Facsimile **and** Transcription
- Not character codes, but position of the characters in the text as primary reference
- Multiple representations can then be related positional

Conceptual view of the model

Figure: Layers of text stacked upon each other

- input according to the Chongkan edition

Establishing a workflow for the DZJY edition

- input according to the Chongkan edition
- scan of page images, image preparation

Establishing a workflow for the DZJY edition

- input according to the Chongkan edition
- scan of page images, image preparation
- upload into the editing environment

Establishing a workflow for the DZJY edition

- input according to the Chongkan edition
- scan of page images, image preparation
- upload into the editing environment
- proofreading and annotating

Establishing a workflow for the DZJY edition

- input according to the Chongkan edition
- scan of page images, image preparation
- upload into the editing environment
- proofreading and annotating
- workbench for analysis, annotation and translation

The editing environment

Figure: A page from the DZJY Chongkan edition in the editing environment

.. in action

Requirements for a scholarly digital edition

- establishing of a reliable text

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile
- user/reader has full control over (his version of) the text

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile
- user/reader has full control over (his version of) the text
- annotation, commentary, translation

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile
- user/reader has full control over (his version of) the text
- annotation, commentary, translation
    - private, shared or public

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile
- user/reader has full control over (his version of) the text
- annotation, commentary, translation
    - private, shared or public
    - the user retains control over his work

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile
- user/reader has full control over (his version of) the text
- annotation, commentary, translation
    - private, shared or public
    - the user retains control over his work
- The pedigree of the edition needs to be explicit

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile
- user/reader has full control over (his version of) the text
- annotation, commentary, translation
  - private, shared or public
  - the user retains control over his work
- The pedigree of the edition needs to be explicit
  - for all public and private versions, there needs to be a way to establish the differences and the degree of relationship

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile
- user/reader has full control over (his version of) the text
- annotation, commentary, translation
    - private, shared or public
    - the user retains control over his work
- The pedigree of the edition needs to be explicit
    - for all public and private versions, there needs to be a way to establish the differences and the degree of relationship
    - it should be possible to

Requirements for a scholarly digital edition

- establishing of a reliable text
- all major witnesses available as facsimile
- user/reader has full control over (his version of) the text
- annotation, commentary, translation
    - private, shared or public
    - the user retains control over his work
- The pedigree of the edition needs to be explicit
    - for all public and private versions, there needs to be a way to establish the differences and the degree of relationship
    - it should be possible to
- Distributed version control systems to the rescue?

- DVCS evolved out of traditional (centralized) version control systems

Distributed version control systems (DVCS)

- DVCS evolved out of traditional (centralized) version control systems
- DVCS do not require a central server or central control

Distributed version control systems (DVCS)

- DVCS evolved out of traditional (centralized) version control systems
- DVCS do not require a central server or central control
- users can freely create their own branches, but retain the connection to the master repository

Distributed version control systems (DVCS)

- DVCS evolved out of traditional (centralized) version control systems
- DVCS do not require a central server or central control
- users can freely create their own branches, but retain the connection to the master repository
- Quick and wide adaption in the open source community

Distributed version control systems (DVCS)

- DVCS evolved out of traditional (centralized) version control systems
- DVCS do not require a central server or central control
- users can freely create their own branches, but retain the connection to the master repository
- Quick and wide adaption in the open source community
- seems ideal suited for the research community

Distributed version control systems (DVCS)

- DVCS evolved out of traditional (centralized) version control systems
- DVCS do not require a central server or central control
- users can freely create their own branches, but retain the connection to the master repository
- Quick and wide adaption in the open source community
- seems ideal suited for the research community
  - but there might be barriers to the entry

Example of a DVCS based workflow

Git Data Transport Commands
http://osteele.com
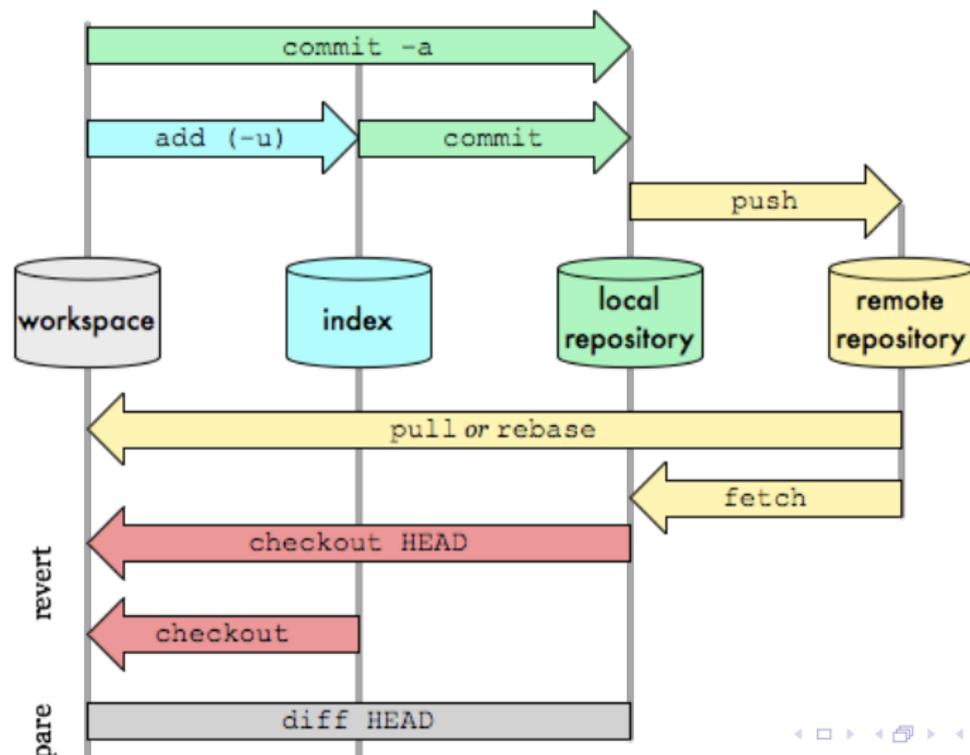
Example(1): Zutang ji (Record of the Hall of Patriarchs)

Figure: The established text of the ZTJ

Example(2) ZTJ text with translation

```
<pb:ZTJ_K45n1503_1-101>¶
祖堂集卷第三¶        Zǔtáng jí Fascicle Three
¶
牛頭和尚,              Preceptor Niútóu
嗣四祖。              succeeded to the Fourth Patriarch;
師諱法融,              his canonical name as master was Fǎróng;
潤州延陵人也,          he was a man from Yánlíng in Rùn province
姓¶文。               and his family name was Wén.
四祖在雙峰山, 告眾曰:      The Fourth Patriarch addressed the
U:---ztj.org           11% (1591,0)  Git:translation  (
```

Figure: ZTJ text, reformatted and with translation (by C. Anderl)

Conclusions

- A new way of handling for digital Chinese texts has been proposed

Conclusions

- A new way of handling for digital Chinese texts has been proposed
- Some changes to the way text is established

Conclusions

- A new way of handling for digital Chinese texts has been proposed
- Some changes to the way text is established
- A different model for publications of a text

Conclusions

- A new way of handling for digital Chinese texts has been proposed

- Some changes to the way text is established

- A different model for publications of a text

- Empower the readers to share as much (or as little) of the traces of their reading with as many (or as few) others

Conclusions

- A new way of handling for digital Chinese texts has been proposed

- Some changes to the way text is established

- A different model for publications of a text

- Empower the readers to share as much (or as little) of the traces of their reading with as many (or as few) others

- Work in progress, update to follow