Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○○○○○○○○○○○

New text format
○○○○○

Open Questions
○○○

Next steps for the TLS Website

Christian Wittern

Kyoto University Institute for Research in Humanities Center for Informatics in East-Asian Studies

2022-01-28

Next steps for the TLS Website

Overview
○

The Kanseki Repository
●○○

Commentaries and textual variants
○○○○○○○○○○○

New text format
○○○○○

Open Questions
○○○

▶ 3 years of development on the 3rd version since Filemaker

▶ the core of the internal structure still reflects a mind-frame of relational database

▶ time to learn from these experiences and re-implement the site

▶ back to the drawing board:

    ▶ (internal) represenation of a text

    ▶ represenation of related texts

    ▶ connection to other parts

Next steps for the TLS Website

KR Design considerations

- We think of a text as represented through the different documents that carried the text throughout the ages
- We want to study the differences we see between the different documents that carry the text.
- These are part of the record , our critical edition is a new interpretation , which considers all relevant information
- As a system, it allows users to retain control over their work

KR System overview

- ▶ The KR holds ca 9800 texts in 6 main categories and 81 subcategories
- ▶ All the texts are available publicly at https://github.com/kanripo
    - ▶ editions (documentary or interpretative) are represented as branches
- ▶ The texts can be searched and read at https://www.kanripo.org
- ▶ An API is available for integration with other projects at https://www.kanripo.org/api/v1.0/search
- ▶ Mandoku, an Emacs Module can be used to search and read the texts
- ▶ Publicly available since March 2016

Next steps for the TLS Website

Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○●○○○○○○○○○

New text format
○○○○○

Open Questions
○○○

Text and Commentary: (1)

Text and Commentary: (2)

KanripoX

- ▶ A new tagset with three modules:
    - ▶ Manifest
        - ▶ itemized list of related materials and metadata
    - ▶ Token
        - ▶ A new simplified "shadow" of the text files
    - ▶ Nexus
        - ▶ List of links between each of the text files
- ▶ see KanripoX: A tagset for connecting digital texts

Manifest

- ▶ grouping and selecting of texts
- ▶ this also includes identifying and relating parts of texts in a regular way, such as a root text to commentaries or translations.
- ▶ It is also be a convenient way to maintain metadata related to the texts.
- ▶ The implementation should allow users to create their own manifest files, which will make it possible to make collections and metadata adjustable.
- ▶ main tags: `<editionGroup>`, `<edition>`, `<divisions>`

Manifest example

```
1  <manifest xmlns="http://kanripo.org/ns/KRX/1.0" xmlns:tei="http://www.tei-c.org/ns/1.0"
2      xml:id="KR5c0057">
3      <description>Text and commentaries of the Daodejing</description>
4      <editions>
5          <editionGroup type="root">
6              <creation>
7                  <date notbefore="-300" notafter="-200" cert="middle">Middle of 3th century BC</date>
8                  <resp role="author">Lao Er老耳, alleged Family name 李, courtesy name 聃.</resp>
9              </creation>
10             <edition id="KR5c0057_tls" format="xml/TEI" location="int/KR5c0057_tls">
18             <edition id="KX5c0045_master" format="xml/TEI" location="int/KX5c0045_master">
30             <edition id="KX5c0045_HFL" format="xml/TEI" location="doc/KX5c0045_HFL">
34             <edition id="KX5c0045_ZTDZ" format="xml/TEI" location="doc/KX5c0045_ZTDZ">
35                 type="documentary" language="lzh" role="reference">
36                 <description>道德真經(正文)</description>
37             </edition>
38             <edition id="CH1a0918a_chant" format="txt/mandoku" location="int/CH1a0918a_chant">
45             <edition id="CH1a0918a_other" format="txt/mandoku" location="int/CH1a0918a_other">
46                 type="interpretative" language="lzh">
47                 <description>老子</description>
48                 <tokenmap>
49                     <map src="^[\*]* " tok="h"/>
50                 </tokenmap>
51             </edition>
52             <edition id="CH8x3004_chant" format="txt/mandoku" location="doc/CH8x3004_chant">
53                 type="documentary" language="lzh">
54                 <description>老子甲本</description>
55                 <divisions>
56                     <div sequence="2" label="德經" start="0" end="3034"/>
57                     <div sequence="1" label="道經" start="3035" end="5432"/>
58                 </divisions>
```

Token

- ▶ Every text is converted to a list of tokens (`<t>`)
- ▶ Tokens can be grouped together as token groups (`<tg>`)
- ▶ required attributes:
  - ▶ @tp: token position
  - ▶ @role: (part of heading, paragraph, note . . .)
  - ▶ @n: label or identifier in the text
- ▶ optional attributes (selection):
  - ▶ @f, @p: (f)ollowing or (p)receding punctuation
  - ▶ @cp: codepoint of character appearing in text

Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○○○○○○○○●○○○

New text format
○○○○○

Open Questions
○○○

Token example

```xml
1  <?xml version="1.0" encoding="UTF-8"?>
2  <tList xml:id="KR5c0057_tls-tok-0000" ed="KR5c0057_tls" n="tok-0000"
3     xmlns="http://kanripo.org/ns/KRX/1.0">
4     <tg>
5        <t tp="0" n="KR5c0057_tls_001-1a.3-h" role="h" pos="1">第</t>
6        <t tp="1" n="KR5c0057_tls_001-1a.3-h" role="h" pos="2">一</t>
7        <t tp="2" n="KR5c0057_tls_001-1a.3-h" role="h" pos="3">章</t>
8     </tg>
9     <tg xml:id="KR5c0057_tls_001-1a.3">
10       <lb ed="KR5c0057_tls" n="KR5c0057_tls_001-1a.3"/>
11       <t tp="3" n="KR5c0057_tls_001-1a.3" role="p" pos="1">道</t>
12       <t tp="4" n="KR5c0057_tls_001-1a.3" role="p" pos="2">可</t>
13       <t tp="5" n="KR5c0057_tls_001-1a.3" role="p" pos="3" f="，">道</t>
14    </tg>
15    <tg xml:id="KR5c0057_tls_001-1a.4">
16       <lb ed="KR5c0057_tls" n="KR5c0057_tls_001-1a.4"/>
17       <t tp="6" n="KR5c0057_tls_001-1a.4" role="p" pos="1">非</t>
18       <t tp="7" n="KR5c0057_tls_001-1a.4" role="p" pos="2">恒</t>
19       <t tp="8" n="KR5c0057_tls_001-1a.4" role="p" pos="3" f="。">道</t>
20    </tg>
21    <tg xml:id="KR5c0057_tls_001-1a.5">
22       <lb ed="KR5c0057_tls" n="KR5c0057_tls_001-1a.5"/>
23       <t tp="9" n="KR5c0057_tls_001-1a.5" role="p" pos="1">名</t>
24       <t tp="10" n="KR5c0057_tls_001-1a.5" role="p" pos="2">可</t>
25       <t tp="11" n="KR5c0057_tls_001-1a.5" role="p" pos="3" f="，">名</t>
26    </tg>
27    <tg xml:id="KR5c0057_tls_001-1a.6">
28       <lb ed="KR5c0057_tls" n="KR5c0057_tls_001-1a.6"/>
29       <t tp="12" n="KR5c0057_tls_001-1a.6" role="p" pos="1">非</t>
30       <t tp="13" n="KR5c0057_tls_001-1a.6" role="p" pos="2">恒</t>
31       <t tp="14" n="KR5c0057_tls_001-1a.6" role="p" pos="3" f="。">名</t>
```

Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○○○○○○○○○●○○

New text format
○○○○○

Open Questions
○○○

Nexus

- For every `<edition>` in the manifest file, there should be a nexus file
- This file gives text locations in the other files for every text
- A nexus file consists of a `<nexusList>` with any number of `<nexus>` elements
- These will have any number of `<locationRef>` elements, that point to text locations
- required elements are @ed, @target, @tp
- @target points to the text element identifier, @tp gives the token position within the text

Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○○○○○○○○○○●○

New text format
○○○○○

Open Questions
○○○

Nexus example

```xml
<nexus xml:id='KR5c0057_tls_001-1a.5' tp='9' tcount='3'>
<locationRef ed='CH1a0918a_chant' target='CH1a0918_CHANT_001-1a.4' tp='8' tcount='3'/>
<locationRef ed='CH1a0918a_other' target='CH1a0918_CHANT_001-1a.4' tp='14' tcount='4'/>
<locationRef ed='CH1a0918b_chant' target='CH1a0918_CHANT_082-2a.11' tp='57' tcount='17'/>
<locationRef ed='CH1a0918b_other' target='CH1a0918_CHANT_082-2a.10' tp='51' tcount='14'/>
<locationRef ed='CH8x3004_chant' target='CH8x3004_CHANT_002-1a.11' tp='3055' tcount='4'/>
<locationRef ed='CH8x3005_chant' target='CH8x3005_CHANT_001-15a.17' tp='4530' tcount='3'/>
<locationRef ed='CH8x3006_chant' target='CH8x3006_CHANT_001-1a.190' tp='384' tcount='3'/>
<locationRef ed='CH8x3007_chant' target='CH8x3007_CHANT_002-1a.13' tp='3052' tcount='4'/>
<locationRef ed='KR5c0073_tls' target='KR5c0073_tls.seg1-3' tp='5' tcount='3'/>
<locationRef ed='KX5c0045_HFL' target='KX5c0045_HFL_001-001a.03' tp='18' tcount='3'/>
<locationRef ed='KX5c0045_ZTDZ' target='KX5c0045_SJB_001-110474b.03' tp='18' tcount='3'/>
<locationRef ed='KX5c0046_HFL' target='KX5c0046_HFL_000-001a.03' tp='21' tcount='3'/>
<locationRef ed='KX5c0046_ZTDZ' target='KX5c0046_SJB_000-110482a.03' tp='21' tcount='3'/>
<locationRef ed='KX5c0065_HFL' target='.13' tp='54' tcount='13'/>
<locationRef ed='KX5c0065_SBCK' target='KX5c0065_SBCK_001-1a.05' tp='964' tcount='13'/>
<locationRef ed='KX5c0065_ZTDZ' target='KX5c0065_SJB_001-120001a.06' tp='64' tcount='13'/>
<locationRef ed='KX5c0073_HFL' target='KX5c0073_HFL_001-001a.03' tp='20' tcount='3'/>
<locationRef ed='KX5c0073_ZTDZ' target='KX5c0073_SJB_001-120272c.03' tp='20' tcount='3'/>
</nexus>
<nexus xml:id='KR5c0057_tls_001-1a.6' tp='12' tcount='3'>
<locationRef ed='CH1a0918a_chant' target='CH1a0918_CHANT_001-1a.5' tp='11' tcount='3'/>
<locationRef ed='CH1a0918a_other' target='CH1a0918_CHANT_001-1a.5' tp='18' tcount='8'/>
<locationRef ed='CH1a0918b_chant' target='CH1a0918_CHANT_082-3a.6' tp='74' tcount='47'/>
<locationRef ed='CH1a0918b_other' target='CH1a0918_CHANT_082-3a.5' tp='65' tcount='54'/>
<locationRef ed='CH8x3004_chant' target='CH8x3004_CHANT_002-1a.13' tp='3059' tcount='3'/>
<locationRef ed='CH8x3005_chant' target='CH8x3005_CHANT_002-1a.44' tp='5069' tcount='3'/>
<locationRef ed='CH8x3006_chant' target='CH8x3006_CHANT_001-1a.190' tp='386' tcount='3'/>
<locationRef ed='CH8x3007_chant' target='CH8x3007_CHANT_002-1a.17' tp='3056' tcount='3'/>
<locationRef ed='KR5c0073_tls' target='KR5c0073_tls.seg1-4' tp='8' tcount='27'/>
<locationRef ed='KX5c0045_HFL' target='KX5c0045_HFL_001-001a.03' tp='21' tcount='3'/>
<locationRef ed='KX5c0045_ZTDZ' target='KX5c0045_SJB_001-110474b.03' tp='21' tcount='3'/>
<locationRef ed='KX5c0046_HFL' target='KX5c0046_HFL_000-001a.03' tp='24' tcount='3'/>
<locationRef ed='KX5c0046_ZTDZ' target='KX5c0046_SJB_000-110482a.03' tp='24' tcount='3'/>
<locationRef ed='KX5c0065_HFL' target='.14' tp='67' tcount='43'/>
```

Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○○○○○○○○○○●

New text format
○○○○○

Open Questions
○○○

KanripoX in action

Figure: Screenshot of Laozi 1 on HXWD

Next steps for the TLS Website

Overview

The Kanseki Repository

Commentaries and textual variants

New text format

Open Questions

Aims

- ▶ Align related texts also within a repository
  - ▶ text & commentary
- ▶ Support critical editing of the text
- ▶ Allow for text analytical processing
- ▶

Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○○○○○○○○○○○

New text format
○○●○○

Open Questions
○○○

Old format and new format #1


```
<div>
    <head>
        <seg xml:id="KR5c0057_tls_001-1a.3-h">1 第一章</seg>
    </head>
    <p>
    </p>
    <p>
        <seg xml:id="KR5c0057_tls_001-1a.3">道可道，<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.4">非恒道。<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.5">名可名，<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.6">非恒名。<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.7">無名、天地之始，<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.8">有名、萬物之母，<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.9">故恒無欲，以觀其妙。<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.10">恒有欲，以觀其徼。<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.11">此兩者同出而異名。<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.12">同謂之玄。<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.13">玄之又玄，<lb/>
        </seg>
        <seg xml:id="KR5c0057_tls_001-1a.14">眾妙之門。</seg>
    </p>
    <p>
        <seg xml:id="KR5c0057_tls_002-1a.1">
            <pb xml:id="KR5c0057_tls_002-1a" ed="tls" n="002-1a"/>
            <lb/>
        </seg>
    </p>
</div>
```

Old format

```
<div>
<head xml:id="KR5c0057_tls_001-1a.3-h" n="1">第一章</head>
<p xml:id="KR5c0057_tls_p2">
<c/>道可道<c xml:id="KR5c0057_tls_001-1a.3" n="，"/>
非恒道<c xml:id="KR5c0057_tls_001-1a.4" n="。"/>
名可名<c xml:id="KR5c0057_tls_001-1a.5" n="，"/>
非恒名<c xml:id="KR5c0057_tls_001-1a.6" n="。"/>
無名<c n="、" />天地之始<c xml:id="KR5c0057_tls_001-1a.7" n="，"/>
有名<c n="、" />萬物之母<c xml:id="KR5c0057_tls_001-1a.8" n="，"/>
故恒無欲<c n="，" />以觀其妙<c xml:id="KR5c0057_tls_001-1a.9" n="。"/>
恒有欲<c n="，" />以觀其徼<c xml:id="KR5c0057_tls_001-1a.10" n="。"/>
此兩者同出而異名<c xml:id="KR5c0057_tls_001-1a.11" n="。"/>
同謂之玄<c xml:id="KR5c0057_tls_001-1a.12" n="。"/>
玄之又玄<c xml:id="KR5c0057_tls_001-1a.13" n="，"/>
眾妙之門<c xml:id="KR5c0057_tls_001-1a.14" n="。"/>
</p></div>
```

New format #1

Old format and new format #2

```xml
        <div>
            <head>
                <seg xml:id="KR5c0057_tls_001-1a.3-h">1 第一章</seg>
            </head>
            <p>
    </p>
            <p>
                <seg xml:id="KR5c0057_tls_001-1a.3">道可道, <lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.4">非恒道。<lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.5">名可名, <lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.6">非恒名。<lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.7">無名、天地之始, <lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.8">有名、萬物之母, <lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.9">故恒無欲, 以觀其妙。<lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.10">恒有欲, 以觀其徼。<lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.11">此兩者同出而異名。<lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.12">同謂之玄。<lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.13">玄之又玄, <lb/>
                </seg>
                <seg xml:id="KR5c0057_tls_001-1a.14">眾妙之門。</seg>
            </p>
            <p>
                <seg xml:id="KR5c0057_tls_002-1a.1">
                    <pb xml:id="KR5c0057_tls_002-1a" ed="tls" n="002-1a"/>
                    <lb/>
                </seg>
            </p>
```

Old format

```xml
            <div>
<head xml:id="KR5c0057_tls_001-1a.3-h" n="1">第一章</head>
            <p>
<lb xml:id="KR5c0057_tls_001-1a.3" ed="tls"/>道可道<c n=", "/>
<lb xml:id="KR5c0057_tls_001-1a.4" ed="tls"/>非恒道<c n="。"/>
<lb xml:id="KR5c0057_tls_001-1a.5" ed="tls"/>名可名<c n=", "/>
<lb xml:id="KR5c0057_tls_001-1a.6" ed="tls"/>非恒名<c n="。"/>
<lb xml:id="KR5c0057_tls_001-1a.7" ed="tls"/>無名<c n=", "/>天地之始<c n=", "/>
<lb xml:id="KR5c0057_tls_001-1a.8" ed="tls"/>有名<c n=", "/>萬物之母<c n=", "/>
<lb xml:id="KR5c0057_tls_001-1a.9" ed="tls"/>故恒無欲<c n=", "/>以觀其妙<c n="。"/>
<lb xml:id="KR5c0057_tls_001-1a.10" ed="tls"/>恒有欲<c n=", "/>以觀其徼<c n="。"/>
<lb xml:id="KR5c0057_tls_001-1a.11" ed="tls"/>此兩者同出而異名<c n="。"/>
<lb xml:id="KR5c0057_tls_001-1a.12" ed="tls"/>同謂之玄<c n="。"/>
<lb xml:id="KR5c0057_tls_001-1a.13" ed="tls"/>玄之又玄<c n=", "/>
<lb xml:id="KR5c0057_tls_001-1a.14" ed="tls"/>眾妙之門<c n="。"/><lb ed="tls"/></p>
                    <pb xml:id="KR5c0057_tls_002-1a" ed="tls" n="002-1a"/>
                </div>
```

New format #2

Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○○○○○○○○○○○

New text format
○○○○●

Open Questions
○○○

New version side by side

Figure: Compare the two directly

Next steps for the TLS Website

Overview
○

The Kanseki Repository
○○○

Commentaries and textual variants
○○○○○○○○○○○

New text format
○○○○○

Open Questions
○●○

Does the TLS creates a new edition of the texts?

▶ The Chinese text should be established according to the understanding of the reader/translator
▶ Do we need one canonical, authorative version of the text?
▶ Do we relegate text critical notes to the apparatus of the translation?
▶ What position have other versions of the text
    ▶ documentary editions: reference only
    ▶ some might be in facsimile only
    ▶ possible target for search/annotation/translation?

Requirements for the new system

- ▶ Improved search across the text
- ▶ Facsimile of many more editions
- ▶ Workflow for establishing text
- ▶ All Kanripo texts possible target for annotation
- ▶ Stable identifiers for all entities
- ▶ Standard-based, fit for long-term maintenance
  - ▶ Linked Open Data integration, API
- ▶ Maybe based on tei-publisher
- ▶ Other items?