Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	000000		00

# Current state and future plans for the TLS research platform

Christian Wittern

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

2022-05-13

ъ

Kyoto University

Christian Wittern

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
•				

#### Overview

TEI Publisher

Better full text search

Linking lexicon and text

First impression of TLS-TP

▲□ > ▲□ > ▲目 > ▲目 > ■ のへで

Kyoto University

Christian Wittern



Next version will be based on TEI Publisher

Some examples of applictions developed with TEI Publisher

Better fulltext search (requires big update to text format)

Needs some careful considerations

First experiments with TLS on TEI Publisher

Christian Wittern

Current state and future plans for the TLS research platform

• • • • • • • • • • • •

#### Overview

#### **TEI** Publisher

Better full text search

Linking lexicon and text

First impression of TLS-TP

▲□ > ▲□ > ▲目 > ▲目 > ■ のへで

Kyoto University

Christian Wittern

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0●00	000000		00

## About TEI Publisher

- Developed since 2015, current version 7.1.0
- Abstracted layer of implementation for processing TEI texts
- Latest developments:
  - Backend accessed through standardized API
  - Frontend based on custom Webcomponents
  - Interface for adding annotations

Christian Wittern

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
0	00●0	000000		00

## Van Gogh Letter

- Based on TEI Publisher 5.0
- Flexible layout with text, translation, facsimile and notes
- The layout can be configured freely by the user
- Search results include facets
- Original and translation on equal footing

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	000●	000000		00

## Alfred Escher Briefedition

- Based on TEI Publisher 8.0 (not yet released)
- Additional information about related persons and events
- Context of letter exchanges
- Edited text and diplomatic transcription
  - Text of the letter + registry of persons/places/contexts
  - Line-wise link between text and facsimile
- Timeline

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	●00000		00

Overview

TEI Publisher

Better full text search

Linking lexicon and text

First impression of TLS-TP

◆□ > ◆□ > ◆三 > ◆三 > 三 の < ⊙

Kyoto University

Christian Wittern

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	00000		00

## Reminder of the discussion

- Currently, the search function can not search across punctuation
- Every text line (phrase) is treated as an atomic unit
  - this is part of the heritage of the Filemaker database
- To change this, fundamental changes to the text format are required
- A first version of this was proposed in January

イロト イボト イヨト イヨ

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	000000		00

#### Format as discussed in January

	<div></div>
Need Xmt.11d="KR5c005]_[[3_00]+4.3-71 [m 1 2m] 一年(Need- (本) Xml:11d="KR5c005]_[13_02] / / /	<head n="1" xml:id="KR5c0057_tls_001-la.3-h">第一聲</head> <lb ed="tls" xml:id="KR5c0057_tls_001-la.3"></lb> 道可道 <c n=", "></c> <lb ed="tls" xml:id="KR5c0057_tls_001-la.4"></lb> 道可道 <c n=", "></c>
石川石々と XML11d="KR50097_115_001-14.5 nm", /> 非恒名々と XML11d="KR50097_115_001-1a.6 nm", "/> 無名々c nm", "/>天規之始々と XML13d="KR50097_115_001-1a.8" nm", "/> 放恒無欲々c nm", "/>規規之々e XML13d="KR50097_115_001-1a.8" nm", "/> 放恒無欲々c nm", "/>以規規妙々c XML13d="KR50097_115_001-1a.9" nm", "/>	<lb ed="tls" xml:id="KR500957_tls_001:1a.5"></lb> 名可茶cc n=", "/> <lb ed="tls" xml:id="KR500957_tls_001:1a.6"></lb> 指道名cc n=", "/> <lb ed="tls" xml:id="KR500957_tls_001:1a.7"></lb> 無名cc n=", "/>天地之始 <lb ed="tls" xml:id="KR500957_tls_001:1a.8"></lb> 有名cc n=", "/>萬物之母-cc n=", "/> <lb ed="tls" xml:id="KR500957_tls_001:1a.8"></lb> 指気/成田粉(cc n=", "/>萬物之母-cc n=", "/>
此兩者同出而異名 <c n="。" xml:id="KR5c0057_tls_001-1a.11"></c> 同謂之玄 <c n="。" xml:id="KR5c0057_tls_001-1a.12"></c> なン又をc xml:id="KR5c0057_tls_001-1a.13" n="、"/>	Cb Xmlid="KR50097_tls_001-la.10" ed="tls"/>恒有欲cc n=", '/>U觀其徼cc n=", '/> Cb Xmlid="KR50097_tls_001-la.11" ed="tls"/>UmER80HER84cc n=", '/> Cb Xmlid="KR50097_tls_001-la.12" ed="tls"/>问調之茲cc n=", '/> Cb Xmlid="KR50097_tls_001-la.13" ed="tls"/>ZizZizZicc n=", '/> Cb Xmlid="KR50097_tls_001-la.14" ed="tls"/>ZizZiZicc n=", '/> Cb Xmlid="KR50097_tls_001-la.14" ed="tls"/>ZizZiZicc n=", '/> Cb Xmlid="KR50097_tls_001-la.14" ed="tls"/>ZizZiCc n=", '/>Cb ed="tls"/>Cicc n=", '/>
	<pre>clb xmt:10= KR5c0057_tls_001-1a.14 ed= tls /&gt;xx05y_rjkc h= " /&gt;clb ed= tls /&gt;xmt ed= tls /&gt;xmt ed= tls /&gt;kmt ed= tls /&gt;xmt ed= tls //xmt ed= tls //xmt</pre>

#### Figure: Change <seg> to <lb>

Kyoto University

Christian Wittern

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
0	0000	000●00		00

#### **Evaluation**

- The format deviates radically from the current one
  - it completely eliminates <seg>
  - but <seg> is the center node of links to translation and lexicon
- This requires extensive changes to many parts of the database
- Further investigations of the working of TEI Publisher revealed
  - eliminating <seg> is not necessary and not desirable
  - a new format that does continue to use <seg> is a better solution

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	0000●0		00

#### Latest proposal for new text format

- Minimally necessary changes to current format
- keep <seg> for linking to other parts
- Problem: texts without punctuation do not have <seg>
  - Other ways to link?
- Consider embedding link anchors into the text
  - How is linking done actually?

Overview O	TEI Publisher 0000	Better full text search 00000●	Linking lexicon and text	First impression of TLS-TP 00
Now to	xt format			
INEW LE	Xt IOIIIat			
	23 V <div></div>			
	24 <head n="&lt;/th" xml:id="KR5&lt;/th&gt;&lt;th&gt;c0057 tls 001-1a.3-h-h"><th>="1"&gt;<sea< th=""><th></th></sea<></th></head>	="1"> <sea< th=""><th></th></sea<>		
		tls 001-1a.3-h">第一章 <th></th> <th></th>		
	25			
	26 <s< th=""><th>eq xml:id="KR5c0057 tls 00</th><th>01-1a.3"&gt;道可道<c_n=", "=""></c_n=",><!--</th--><th>′seq&gt;<seq< th=""></seq<></th></th></s<>	eq xml:id="KR5c0057 tls 00	01-1a.3">道可道 <c_n=", "=""></c_n=",> </th <th>′seq&gt;<seq< th=""></seq<></th>	′seq> <seq< th=""></seq<>
	<pre>26 xml:id="KR5c0057</pre>	tīs 001-1a.4">非恒道 <c n="。&lt;/th&gt;&lt;th&gt;"></c> <seq< th=""><th>5 5</th></seq<>	5 5	
	<pre>20 xml:id="KR5c0057"</pre>	tls_001-1a.5">名可名 <c n="&lt;/th&gt;&lt;th&gt;, "></c> <seg< th=""><th></th></seg<>		
	<pre>20 xml:id="KR5c0057"</pre>	tls_001-1a.6">非恒名 <c n="。&lt;/th&gt;&lt;th&gt;"></c> <seg< th=""><th></th></seg<>		
	<pre>20 xml:id="KR5c0057"</pre>	tls_001-1a.7">無名 <c n="、&lt;/th&gt;&lt;th&gt;"></c> 天地之始 <c_n=","></c_n=",">	<seq< th=""></seq<>	
	<pre>20 xml:id="KR5c0057"</pre>	tls_001-1a.8">有名 <c n="、&lt;/th&gt;&lt;th&gt;"></c> 萬物之母 <c_n=","></c_n=",">	<seg< th=""></seg<>	
	<pre>20 xml:id="KR5c0057"</pre>	tls_001-1a.9">故恒無欲 <c_n=< th=""><th>=","/&gt;以觀其妙<c n="。"></c><th>eg&gt;<seg< th=""></seg<></th></th></c_n=<>	=","/>以觀其妙 <c n="。"></c> <th>eg&gt;<seg< th=""></seg<></th>	eg> <seg< th=""></seg<>
	<pre>20 xml:id="KR5c0057_</pre>	tls_001-1a.10">恒有欲 <c n="&lt;/th"><th>","/&gt;以觀其徼<c n="。"></c><th>eg&gt;<seg< th=""></seg<></th></th></c>	","/>以觀其徼 <c n="。"></c> <th>eg&gt;<seg< th=""></seg<></th>	eg> <seg< th=""></seg<>
		tls_001-1a.11">此兩者同出而		
	<pre>20 xml:id="KR5c0057"</pre>	tls_001-1a.12">同謂之玄 <c r<="" th=""><th>n="。"/&gt;<seg< th=""><th></th></seg<></th></c>	n="。"/> <seg< th=""><th></th></seg<>	
	xml·id-"KP5c0057	+le_001_1a_13"\女力又女~c_r		

- 28 xml:id="KR5c0057\_tls\_001-1a.13">玄之又玄<c n=", "/></seg><seg</pre>
- xml:id="KR5c0057\_tls\_001-1a.14">眾妙之門<c n="。"/></seg>
  xml:id="KR5c0057\_tls\_001-1a.14">眾妙之門<c n="。"/></seg>
  xml:id="KR5c0057\_tls\_002-1a"
  - </div>

28

Figure: Keep <seg>, only add <c>

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	000000	•000000	00

Overview

TEI Publisher

Better full text search

Linking lexicon and text

First impression of TLS-TP

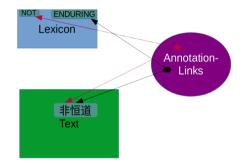
◆□ > ◆□ > ◆三 > ◆三 > 三 の < ⊙

Christian Wittern Current state and future plans for the TLS research platform

Kyoto University

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	000000	○●○○○○○	00

### Linking through annotation file



#### Figure: Text and Lexicon not aware of links

イロン イロン イヨン イヨン

Ξ.

Kyoto University

Christian Wittern

Overview O	TEI Publisher 0000	Better full text search 000000	Linking lexicon and text	First impression of TLS-TP 00
Mechanics		(1) s:tls="http://hxwd.org/ns/1.0" cd ="uuid-2785bda2-fc81-4400-bfbe-5		
		uid-eb80b06c-f691-4d2d-a753-cd180		
			32a30364-5261-438d-8ff8-f493af6e0d	
	<tls::< td=""><td>srcline title="老子" target="#KR5 非恒道。 line title="老子 (en)" src="Karlg</td><td></td><td></td></tls::<>	srcline title="老子" target="#KR5 非恒道。 line title="老子 (en)" src="Karlg		
		ay,	ren 1975 >15 not the constant	
	<pre><form <orth="" cori=""></form></pre>	- resp="#uuid-00e22256-d177-459e-bo ⇒非	-	
	<pron </pron 	<pre>xml:lang="zh-Latn-x-pinyin"&gt;fëi&lt;</pre>		
	< <u>sense</u> co <gram(< td=""><td>rresp="#uuid-32a30364-5261-438d-&amp; Grp&gt;</td><td>3ff8-f493af6e0d17"&gt;</td><td></td></gram(<>	rresp="#uuid-32a30364-5261-438d-& Grp>	3ff8-f493af6e0d17">	
	<	bos>V tls:syn-func		
	ka	<pre>corresp="#uuid-c87f5e8b-6512-4 N</pre>	104d-84b2-9e99a85aa28e">vt	
	<1	<pre>tls:sem-feat     corresp="#uuid-52f9b87c-5688-4     &gt;copula</pre>	1b46-b992-a5fb0bf27fb9"	
		<pre>usg type="warring-states-currency</pre>	/">5	
		nGrp> is not N		▲目▶▲目▶ 目 少へで
Christian Wittern				Kyoto University

# Mechanics of linking (2)

#### </sense>

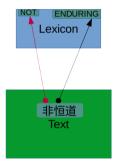
<sense xml:id="uuid-0f7d2e76-ffb9-4604-8776-c0a35cd5c953">

Figure: 非 in concept NOT

Kvoto University

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	000000		00

### Linking without annotation file



#### Figure: Link anchors embedded in text

イロン イロン イヨン イヨン

æ

Kyoto University

Christian Wittern

Overview O	TEI Publisher 0000	Better full text search 000000	Linking lexicon and text 00000●0	First impression of TLS-TP 00		
New ter	t format with	embedded ancho	rc			
		embedded aneno	15			
	23 - <div></div>					
	<pre>24 <head <="" td="" xml:id="KR5c0&lt;/pre&gt;&lt;/td&gt;&lt;td&gt;057_tls_001-1a.3-h-h"><td>"&gt;<seg< td=""><td></td></seg<></td></head></pre>	"> <seg< td=""><td></td></seg<>				
	<pre>24 xml:id="KR5c0057_tl</pre>	s_001-1a.3-h">第一章<	/head>			
	25					
			1a.3">道可道 <c n=", "></c> <s< td=""><td>seg</td></s<>	seg		
		xml:id="KR5c0057_tls_001-1a.4"> <anchor< td=""></anchor<>				
		<pre>xml:id="uuid-eb80b06c-f691-4d2d-a753-cd180a59e2c2"</pre>				
	<pre>&gt;&gt; [ref="uuid-32a30364-5261-438d-8ff8-f493af6e0d17"] resp="#CH"</pre>					
		5T23:05:39.385+09:00" <u>len</u> =				
		8d-5c2f-423b-82eb-b733838c				
		84c8-4b48-bb71-67429471858				
			"1"_type="SWL"/>恒道 <c :<="" n="。" th=""><th>&gt;<seg< th=""></seg<></th></c>	> <seg< th=""></seg<>		
		s_001-1a.5">名可名 <c ,<="" n="," th=""><th></th><th></th></c>				
		s_001-1a.6">非恒名 <c .<="" n="。" th=""><th></th><th></th></c>				
		s_001-1a.7">無名 <c_n="、"></c_n="、">				
		fl-db75-4ce9-931e-43a50aa2				

- 26 ref="uuid-0bbd0bfa-0082-4c14-96e2-86c6437036f5" resp="#CH"
- >>> modified="2019-03-26T13:08:54.554+09:00" len="1" type="SWL"/>始<c n=", "/></seg><seg</pre>
- 20 xml:id="KR5c0057\_tls\_001-1a.8">有名<c n="、"/>萬物之<anchor</pre>
- 20 xml:id="uuid-61a5c49f-f63a-471b-a1f7-29edbc2aa633"
- 26 ref="uuid-e23a5301-ee04-40f6-b693-8d6ebbc34a7e" resp="#CH"
- 20 modified="2019-03-26T13:08:54.554+09:00" len="1" type="SWL"/>母<c n=", "/></seg><seg</pre>
- xml:id="KR5c0057\_tls\_001-1a.9">故恒無欲<c n=", "/>以觀其妙<c n="。"/></seg><seg</pre>

#### Christian Wittern

3 N - 1

= nar

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	000000	000000●	00

### Text with link anchors

30 🗸	
31	<seg xml:id="KR5c0057_tls_001-1a.3">道可道<c n=", "></c></seg>
32 🗸	<pre><seg xml:id="KR5c0057_tls_001-1a.4"><anchor< pre=""></anchor<></seg></pre>
33	xml:id="uuid-eb80b06c-f691-4d2d-a753-cd180a59e2c2"
34	ref="uuid-32a30364-5261-438d-8ff8-f493af6e0d17"_resp="#CH"
35	modified="2019-11-05T23:05:39.385+09:00"_len="1"_type="SWL"/>
	非 <anchor< th=""></anchor<>
36	xml:id="uuid-3a45c08d-5c2f-423b-82eb-b733838ccacb"
37	<u>ref="uuid-4e5c1744-84c8-4b48-bb71-674294718581"</u>
38	modified="2019-11-05T23:04:00.227+09:00"_len="1"_type="SWL"/>恒道 <c< th=""></c<>
38	n="。"/>
39	<seg xml:id="KR5c0057_tls_001-1a.5">名可名<c n=","></c></seg>
40	<seg xml:id="KR5c0057_tls_001-1a.6">非恒名<c n="。"></c></seg>
41 🛩	<seg xml:id="KR5c0057_tls_001-1a.7">無名<c n="、"></c>天地之<anchor< th=""></anchor<></seg>
42	xml:id="uuid-b56779f1-db75-4ce9-931e-43a50aa2b2db"
43	<u>ref="uuid-0bbd0bfa-0082-4c14-96e2-86c6437036f5"</u>
44	modified="2019-03-26T13:08:54.554+09:00"_len="1"_type="SWL"/>始 <c< th=""></c<>
44	n=", "/>

Figure: Anchor for 非 in Laozi 1

Christian Wittern

<ロト < □ ト < □ ト < ≧ ト < ≧ ト ≧ ト ⊃ Q () Kyoto University

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	000000		●0

Overview

**TEI Publisher** 

Better full text search

Linking lexicon and text

First impression of TLS-TP

▲□ > ▲□ > ▲目 > ▲目 > ■ のへで

Kyoto University

Christian Wittern

Overview	TEI Publisher	Better full text search	Linking lexicon and text	First impression of TLS-TP
O	0000	000000		○●

# What is currently working

- new text format can be displayed
  - translations not yet
- anchors not yet visible
- adding of new anchors (annotation)
  - ordering of list for existing definitions?
- other annotations:
  - Persons, places, organizations etc
- text critical markup
- different display modes
- search needs more work

Kyoto University