



Title	Essential Material for Humanities Researchers! Envisioning a New Path in Humanities Research with Text Data and TEI
Author(s)	Vogatza, Dimitra; Voulgaris, Nikolaos; Yoshiga, Natsuko et al.
Citation	
Version Type	VoR
URL	https://doi.org/10.18910/104315
rights	This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Using TEI tags

What are tags?

1. **<xxxxxx>**A block of text or characters**</xxxxxx>**

↑
Start-tag

Element = a part of text enclosed by tags

↑
End-tag

2. **<xxxxxx />**

↑
Empty tag

This tag allows you to assign meaning at a specific position even if there is no text.

Rules of tagging

1. Start-tags and end-tags **must always be used as a pair** when enclosing a block of text.
2. Pairs of start-tag and end-tag may be **nested**.
3. Nested elements **cannot overlap**.



← This is correct



← This is incorrect

Structure of TEI-formatted Data

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
  </teiHeader>
  <text>
    <front>
      <!-- front matter of copy text, if any, goes here -->
    </front>
    <body>
      <!-- body of copy text goes here -->
    </body>
    <back>
      <!-- back matter of copy text, if any, goes here -->
    </back>
  </text>
</TEI>
```

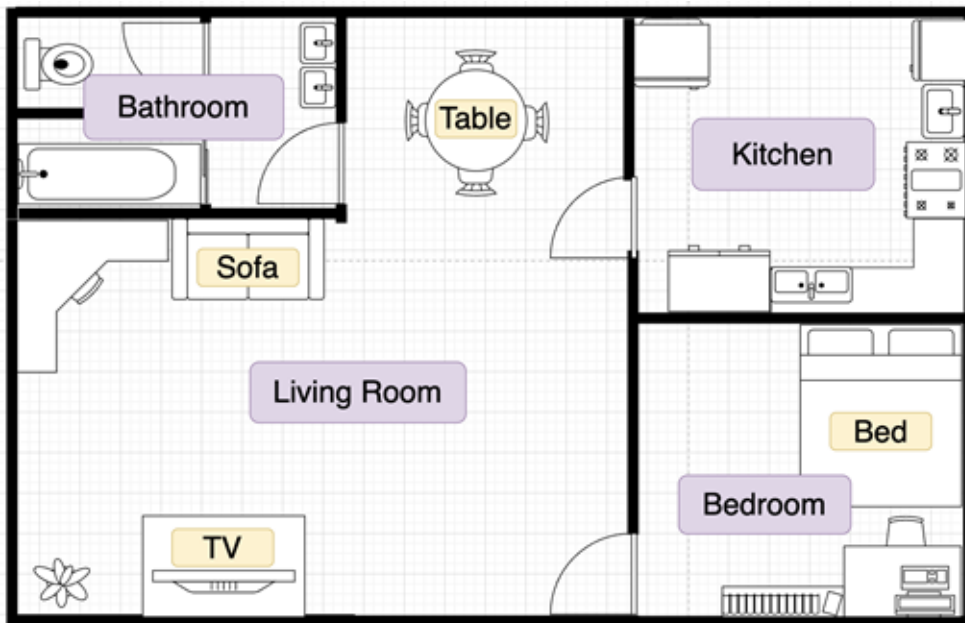
<TEI> Tag (Root Element)

<TEI>

Root Element

(This tag must
always come first)

<TEI>



</TEI>

<teiHeader> Tag

<teiHeader> Tag

A required element that appears inside the <TEI> tag.

The section that manages surrounding information about the text (such as the title, author name, publication date, etc.)

<TEI>

<teiHeader>

Summary or supplementary
information about Main Text

</teiHeader>

<text>

Main Text

</text>

</TEI>

<text> Tag (Main Text)

<text> Tag

Contains the actual content of the text

Structures chapters, paragraphs, headings, etc

<TEI>

<teiHeader>

Summary or supplementary
information about Main Text

</teiHeader>

<text>

Main Text

</text>

</TEI>

Structure of TEI-formatted Data

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
  </teiHeader>
  <text>
    <front>
      <!-- front matter of copy text, if any, goes here -->
    </front>
    <body>
      <!-- body of copy text goes here -->
    </body>
    <back>
      <!-- back matter of copy text, if any, goes here -->
    </back>
  </text>
</TEI>
```

Journal Article (a simple example)

(Bibliography)

Title: My Paper

Author: Taro Yamada

Publisher: UOsaka Publishing

Year of Publication: 2025

(Main Text)

My Paper

Taro Yamada

Introduction

This is the introduction of my paper.

...

<TEI>

<teiHeader>

(Source = Bibliographic
information,
catalog, metadata)

</teiHeader>

<text>

(Main text)

</text>

</TEI>

Tag Structure

(Bibliography)

Title: My Paper

Author: Taro Yamada

Publisher: UOsaka Publishing

Year of Publication: 2025

(Main Text)

My Paper

Taro Yamada

Introduction

This is the introduction of my
paper.

...



```
<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>My Paper</title>
        <author>Taro Yamada</author>
      </titleStmt>
      <publicationStmt>
        <publisher>UOsaka Publishing</publisher>
        <date>2024</date>
      </publicationStmt>
    </fileDesc>
  </teiHeader>
  <text>
    (Main text)
  </text>
</TEI>
```

Using the TEI Guidelines (1)

If you search for “TEI guidelines” in your browser, you’ll find a page like the one on the right: “The TEI Guidelines” (Current Guidelines).

 Guidelines ▾ Activities ▾ Tools ▾ Membership ▾ Support ▾ About ▾ News

Search ▾

TEI: TEIガイドライン

P5 Version 4.8.1. Last updated on 1st November 2024, revision 0a2bff95a

[English] [Deutsch] [Español] [Italiano] [Français] [日本語] [한국어] [中文]



前付

Title

- i. [Releases of the TEI Guidelines](#)
- ii. [Dedication](#)
- iii. [Preface and Acknowledgments](#)
- iv. [About These Guidelines](#)
- v. [A Gentle Introduction to XML](#)
- vi. [Languages and Character Sets](#)

後付

- 付録 A [Model Classes](#)
- 付録 B [Attribute Classes](#)
- 付録 C [Elements](#)
- 付録 D [Attributes](#)
- 付録 E [Datatypes and Other Macros](#)

本文

- 1 [The TEI Infrastructure](#)
- 2 [The TEI Header](#)
- 3 [Elements Available in All TEI Documents](#)
- 4 [Default Text Structure](#)
- 5 [Characters, Glyphs, and Writing Modes](#)
- 6 [Verse](#)
- 7 [Performance Texts](#)
- 8 [Transcriptions of Speech](#)
- 9 [Computer-mediated Communication](#)
- 10 [Dictionaries](#)
- 11 [Manuscript Description](#)
- 12 [Representation of Primary Sources](#)
- 13 [Critical Apparatus](#)
- 14 [Names, Dates, People, and Places](#)

TEI sourcecode

- [Getting and Using the TEI Sources.](#)
- [TEI GitHub Repository.](#)
- [Bug Reports, Feature Requests, etc.](#)

Using the TEI Guidelines (2)

The screenshot shows a web browser interface. At the top, the TEI Guidelines website navigation bar is visible, with the 'teiheader' search term entered in the search box. Below the navigation bar, a large blue arrow points down to a Google search bar. The search bar contains the text 'teiheader site:tei-c.org'. Below the search bar, the search results are displayed. The first result is for the 'Japanese version' of the TEI element 'teiHeader'. The second result is for the 'English version' of the TEI element 'teiHeader'.

TEI: TEIガイドライン

Google

teiheader site:tei-c.org

すべて ショッピング 画像 動画 地図 ニュース 書籍 もっと見る ツール

Text Encoding Initiative
https://tei-c.org › release › doc › tei-p5-doc › html › ref-... : Japanese version

TEI element **teiHeader** (TEI header)

<teiHeader> (TEI header) 全てのTEI準拠テキストに、電子版のタイトルページを構成する、記述的・宣言的情報を示す。 [2.1.1 The TEI

Text Encoding Initiative
https://tei-c.org › release › doc › tei-p5-doc › html › ref-... : English version

TEI element **teiHeader** (TEI header)

<teiHeader>. <teiHeader> (TEI header) supplies descriptive and declarative metadata associated with a digital resource or set of resources. [2.1.1 The ...

Using the TEI Guidelines (3)

<teiHeader>

Home
C Elements

<teiHeader> (TEI header) supplies descriptive and declarative metadata associated with a digital resource or set of resources. [[2.1.1 The TEI Header and Its Components](#) [16.1 Varieties of Composite Text](#)]

Module	header — The TEI Header
Attributes	• att.global: @xml:id, @n, @xml:lang, @xml:base, @xml:space ◦ att.global.analytic: @ana ◦ att.global.change: @change ◦ att.global.facs: @facs ◦ att.global.linking: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select ◦ att.global.rendition: @rend, @style, @rendition ◦ att.global.responsibility: @cert, @resp ◦ att.global.source: @source
Contained by	core: teiCorpus textstructure: TEI
May contain	header: encodingDesc fileDesc profileDesc revisionDesc xenoData
Note	One of the few elements unconditionally required in any TEI document.
Example	<pre><teiHeader> <fileDesc> <titleStmnt> <title>Shakespeare: the first folio (1623) in electronic form</title></pre>

Using the TEI Guidelines (4)

May contain	header: encodingDesc fileDesc profileDesc revisionDesc xenoData
Note	One of the few elements unconditionally required in any TEI document.
Example	<pre><teiHeader> <fileDesc> <titleStmt> <title>Shakespeare: the first folio (1623) in electronic form</title> <author>Shakespeare, William (1564–1616)</author> <respStmt> <resp>Originally prepared by</resp> <name>Trevor Howard-Hill</name> </respStmt> <respStmt> <resp>Revised and edited by</resp> <name>Christine Avern-Carr</name> </respStmt> </titleStmt> <publicationStmt> <distributor>Oxford Text Archive</distributor> <address> <addrLine>13 Banbury Road, Oxford OX2 6NN, UK</addrLine> </address> <idno type="OTA">119</idno> <availability> <p>Freely available on a non-commercial basis.</p> </availability> <date when="1968">1968</date> </publicationStmt> <sourceDesc> <bibl>The first folio of Shakespeare, prepared by Charlton Hinman (The Norton Facsimile, 1968)</bibl> </sourceDesc> </fileDesc> <encodingDesc> <projectDesc> <p>Originally prepared for use in the production of a series of old-spelling concordances in 1968, this text was extensively checked and revised for use during the </projectDesc> </encodingDesc> </teiHeader></pre>  Scroll down and click on the link that says: Show all 

Using the TEI Guidelines (5)

Portion of the screen where all the “teiHeader” examples are displayed

2.1.1 The TEI Header and Its Components

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>
        <!-- title of the resource -->
      </title>
    </titleStmt>
    <publicationStmt>
      <p>
        <!-- Information about distribution of the resource -->
      </p>
    </publicationStmt>
    <sourceDesc>
      <p>
        <!-- Information about source from which the resource derives -->
      </p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

[bibliography](#) ↗

[<teiHeader>](#) (zh-TW)

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Taisho Tripitaka, Electronic version, No. 251 般若波羅蜜多
心經</title>
    <respStmt>
      <resp>底本來源</resp>
      <name>大藏出版株式會社</name>
    </respStmt>
    <respStmt>
      <resp>電子版本製作</resp>
      <name>中華電子佛典協會</name>
    </respStmt>
  </titleStmt>
  <publicationStmt>
    <distributor>中華電子佛典協會</distributor>
    <address>
      <addrLine>11246 台北市北投區光明路276號4樓</addrLine>
    </address>
    <availability>
      <p>本資料使用限定於非營利性用途，並需附上本標頭資訊。</p>
    </availability>
    <date>1998年12月</date>
  </publicationStmt>
  <sourceDesc>
    <bibl>Taisho Tripitaka Vol. 08, Nr. 251 般若波羅蜜多心經</bibl>
  </sourceDesc>
  <encodingDesc>
    <projectDesc>
      <p>預備格式轉換</p>
    </projectDesc>
  </encodingDesc>
  <revisionDesc>
    <list>
      <item>
        <date>19990721/18:35:54</date>CW以CBXML.BAT (99/6/30)轉換成
XML</item>
      </list>
    </revisionDesc>
  </teiHeader>
```

[Show all](#) ↗

Considering the Structure of the Text

(Bibliography)

Title: My Paper

Author: Taro Yamada

Publisher: UOsaka Publishing

Year of Publication: 2025

(Main Text)

My Paper

Taro Yamada

Introduction

This is the introduction of my
paper.

...

1. Understand the overall structure of the text

- Bibliography --- `teiHeader` tag
- Main Text --- `text` tag
 - Chapter
 - Paragraphs
 - Section breaks
 - Annotations
 - Quotations

Starting with the Essential Tags

Let's try expressing the main structure of the text using basic tags

(Bibliography)

Title: My Paper

Author: Taro Yamada

Publisher: UOsaka Publishing

Year of Publication: 2025

(Main Text)

My Paper

Taro Yamada

Introduction

This is the introduction of my paper.

...

Example of a simplified markup

```
<TEI>
  <teiHeader>
    <title>My Paper</title>
    <author>Taro Yamada</author>
    <publisher>UOsaka Publishing</publisher>
    <date>2024</date>
  </teiHeader>
  <text>
    <body>
      Introduction
      This is the introduction of my paper
    </body>
  </text>
</TEI>
```

This is for illustration only and does not represent the complete or accurate TEI structure.

Example of Markup

```
<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>My Paper</title>
        <author>Taro Yamada</author>
      </titleStmt>
      <publicationStmt>
        <distributor>UOsaka Publishing</distributor>
        <date>2024</date>
      </publicationStmt>
      <sourceDesc>
        <p>An example made for this video</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
```



```
    <text>
      <body>
        <div>
          <head>Introduction</head>
          <p>This is the introduction of my paper</p>
        </div>
      </body>
    </text>
  </TEI>
```

(Continues to <text>)

Quick References

This is a recommended site to use as reference for resources in Japanese

上位

textcrit: [lem](#) [rdc](#)
textstructure: [back](#) [body](#) [div](#) [front](#)

下位

analysis: [interp](#) [interpGrp](#) [span](#) [spanGrp](#)
certainty: [certanty](#) [precision](#) [respons](#)
cmc: [post](#)
core: [bibl](#) [biblStruct](#) [cb](#) [cit](#) [desc](#) [divGen](#) [ellips](#)
[milestone](#) [note](#) [noteGrp](#) [p](#) [pb](#) [q](#) [quote](#) [sa](#)
dictionaries: [entry](#) [entryFree](#) [superEntry](#)
drama: [camera](#) [caption](#) [castList](#) [move](#) [sound](#)
figures: [figure](#) [rotatedMusic](#) [table](#)

The bold text indicates **modules**, which are components that group together tags and structures for specific purposes.

“TEI Guidelines Modules Elements, and Attributes List”

https://tei-lab.github.io/element_list/

モジュール一覧

TEIガイドラインのすべての要素は、その用途や使用分野に応じ、以下のモジュールのいずれかに分類されている。モジュール名をクリックすると、要素一覧中の当該モジュールグループの先頭に移動する。

モジュール名	日本語訳モジュール名	ガイドラインの対応する章
tei	全TEIモジュールで使用可能なデータ型、クラス、マクロ。	1. The TEI Infrastructure
header	ヘダーモジュール	2. The TEI Header
core	コアモジュール	3. Elements Available in All TEI Documents
textstructure	テキスト構造モジュール	4. Default Text Structure
gaiji	外字モジュール	5. Characters, Glyphs, and Writing Modes
verse	韻文モジュール	6. Verse
drama	舞台芸術モジュール	7. Performance Texts
translation	翻訳モジュール	8. Translations of Greek

Japanese Resources & Communities on TEI

1. Institute of Humanities and Information Science Research.
Introduction to Text Data Construction for the Humanities: Initiatives Based on the TEI Guidelines.(一般財団法人人文情報学研究所ほか.
人文学のためのテキストデータ構築入門: TEIガイドラインに準拠した取り組みにむけて. 東京, 文学通信) 2022, ISBN978-4-909658-84-5.
2. TEI Research Group <https://tei.dhii.jp/home>
3. Aozora Bunko TEI Automation Project https://github.com/TEI-EAJ/auto_aozora_tei

In Conclusion

Because humanities materials are so diverse, TEI takes the approach that tagging **should be decided by the users or communities handling each type of text.**

So, it's often **fastest and highly recommended** to find TEI examples similar to your text or to try working with hands-on exercises. 