



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

Developing a Research Data Ecosystem for the
Promotion of Data-Driven Science

Research Data Management Teaching Materials –
Practical Edition



The University of Osaka Global Japanese Studies Education and Research

Incubator 「Incubator-Supported Project」

— Developing Digital Humanities Educational Materials to Enhance Research Data
Management in the Humanities —

Essential Material for Humanities Researchers!

Envisioning a New Path in Humanities Research with Text Data and TEI

What is TEI ?

Text Encoding Initiative

TEI (Text Encoding Initiative)

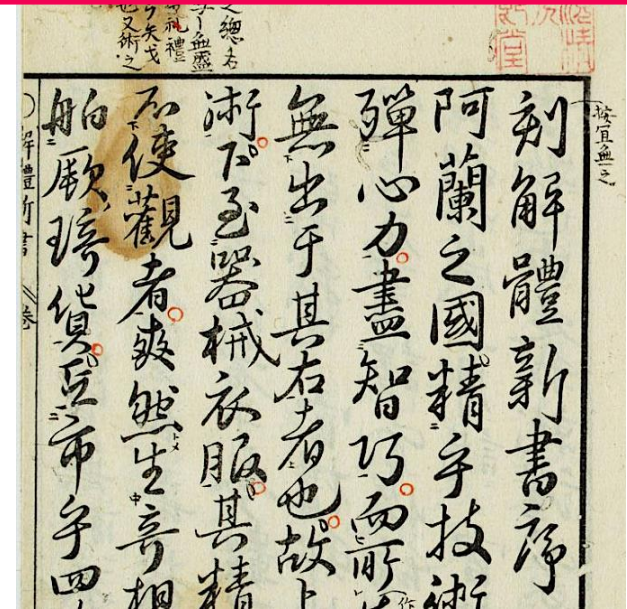
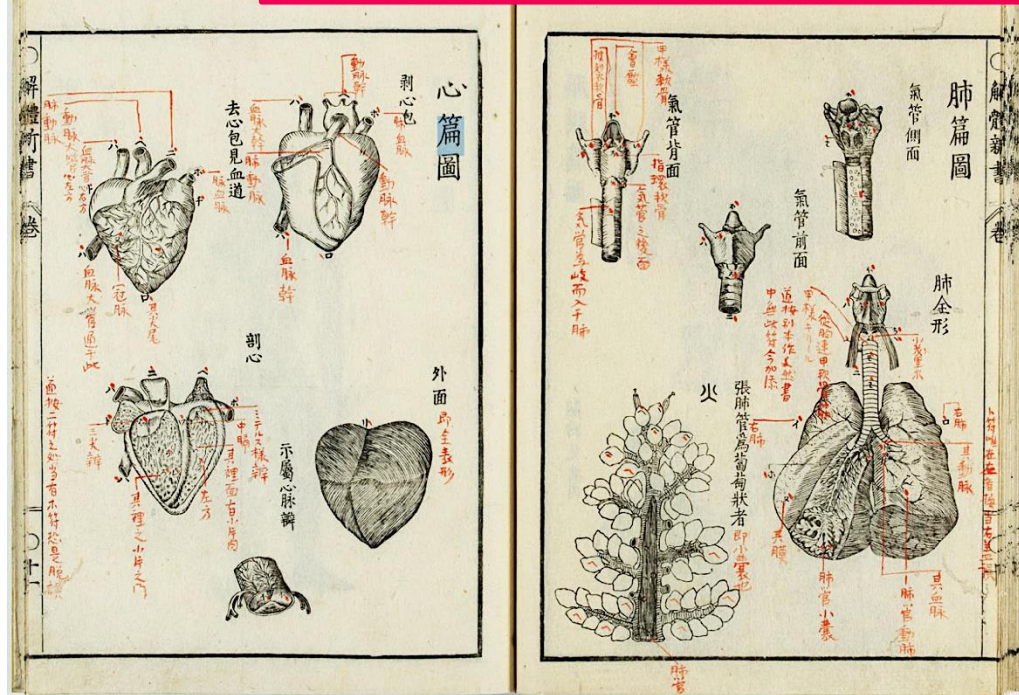
<https://tei-c.org/>

An international organization that compiles guidelines for digitally preserving and representing the **meaning and structure of texts** such as literary works, historical documents, and letters



Purpose and Benefits of TEI (1)

Captures contextual elements beyond the main text—such as bibliographic data, annotations, layout, and seals



“Kaitai Shinsho,” in the collection of the Regional History and Culture Research Center, Saga University

Purpose and Benefits of TEI (2)

XML (Extensible Markup Language)

A method of notation for organizing and structuring the content of text

tag

`<title>The Green Fairy Book</title>`

This part is a title.

I got it.



"The green fairy book" by Lang, Andrew, 1844-1912

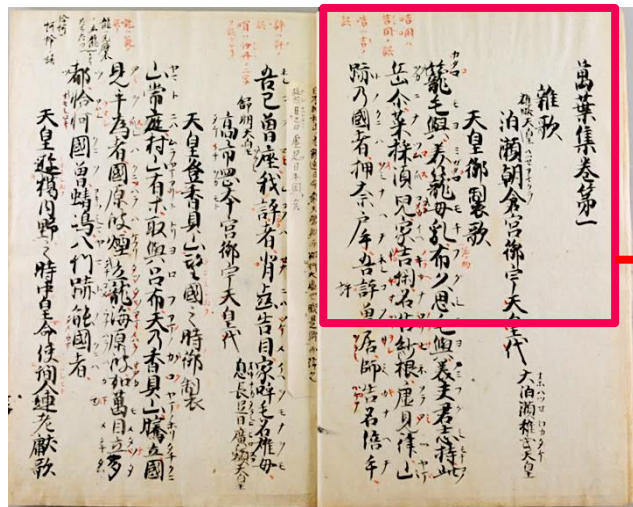
Purpose and Benefits of TEI (3)

Benefits of TEI

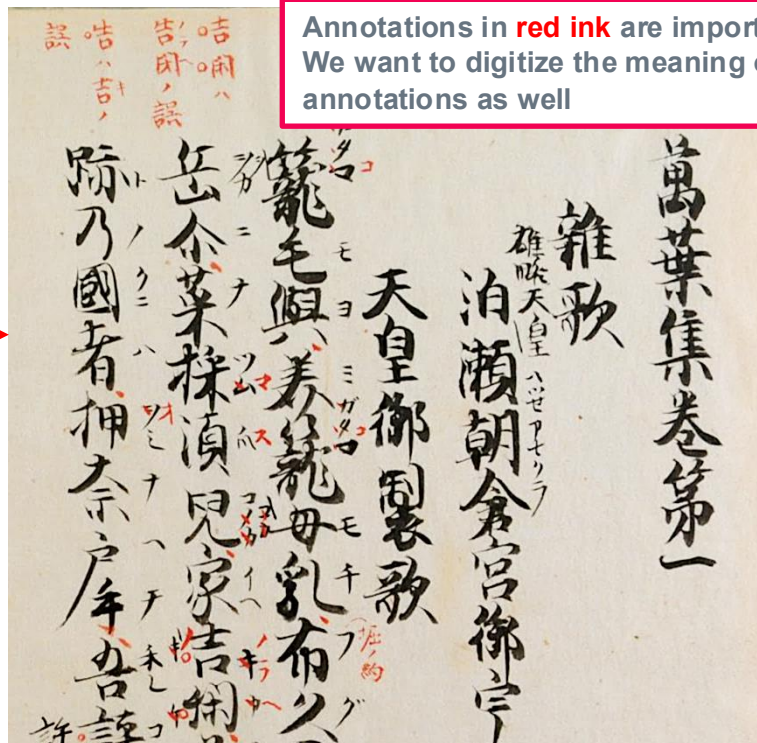
1. **Compatibility** Easy data sharing among researchers
2. **Long-term Preservation** Adaptable to future technological changes
3. **Efficient search and analysis** Makes it easy to find specific information or analyze data based on defined criteria



History and Background of TEI (1)



「萬葉集 20卷」 (Man'yōshū, Volume 20)
Kansai University Asia Open Research
Center [KU-ORCAS]



Annotations in **red ink** are important.
We want to digitize the meaning of the
annotations as well

In 1987, a project began to create a unified digital data format for the humanities → which became TEI

History and Background of TEI (2)

SGML (Standard Generalized Markup Language)

- A technology used primarily in the publishing industry to structure complex documents
- It did not become widely adopted by the general public

The Emergence of **XML** (Extensible Markup Language)

- In the 2000s, XML was introduced as a more user-friendly format for many people
- The **TEI Guidelines (User Guide)** were reconstructed based on XML

History and Background of TEI (3)

Features of XML

1. Tags can be freely defined

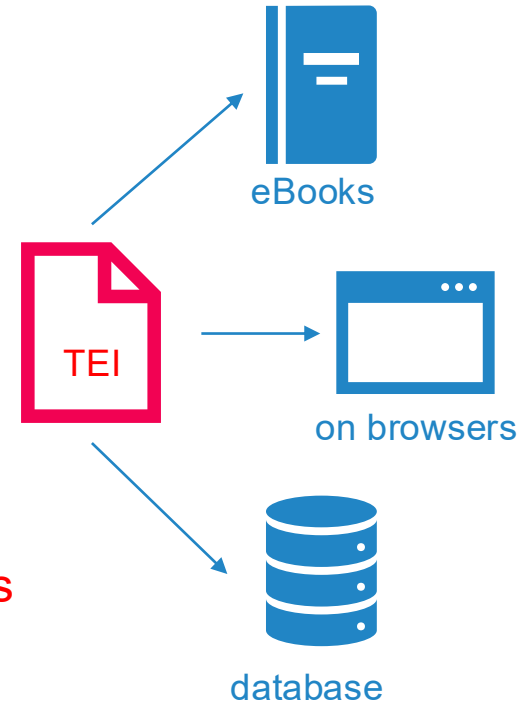
You can define your own tags (labels)

2. Highly extensible

Because tags can be defined to match the structure and content of your data, XML can support a wide variety of data types and use cases.

3. Transformable structure

XML can convert data structure to other formats=
the same data can be reused for various purposes



History and Background of TEI (4)

The TEI's conference and annual general meetings

<https://members.tei-c.org/Events/meetings>



**ENCODING
CULTURES**

JOINT MEC AND TEI
CONFERENCE 2023

TEI Structure (1)

House (floor plan)

- Living Room

-- sofa

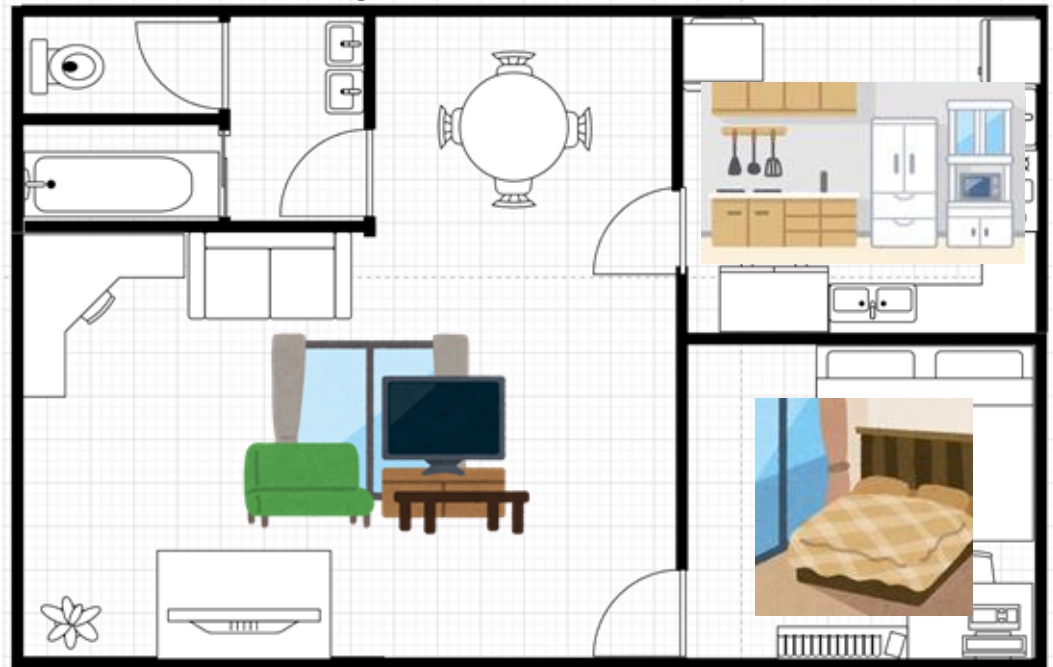
- Kitchen

-- refrigerator

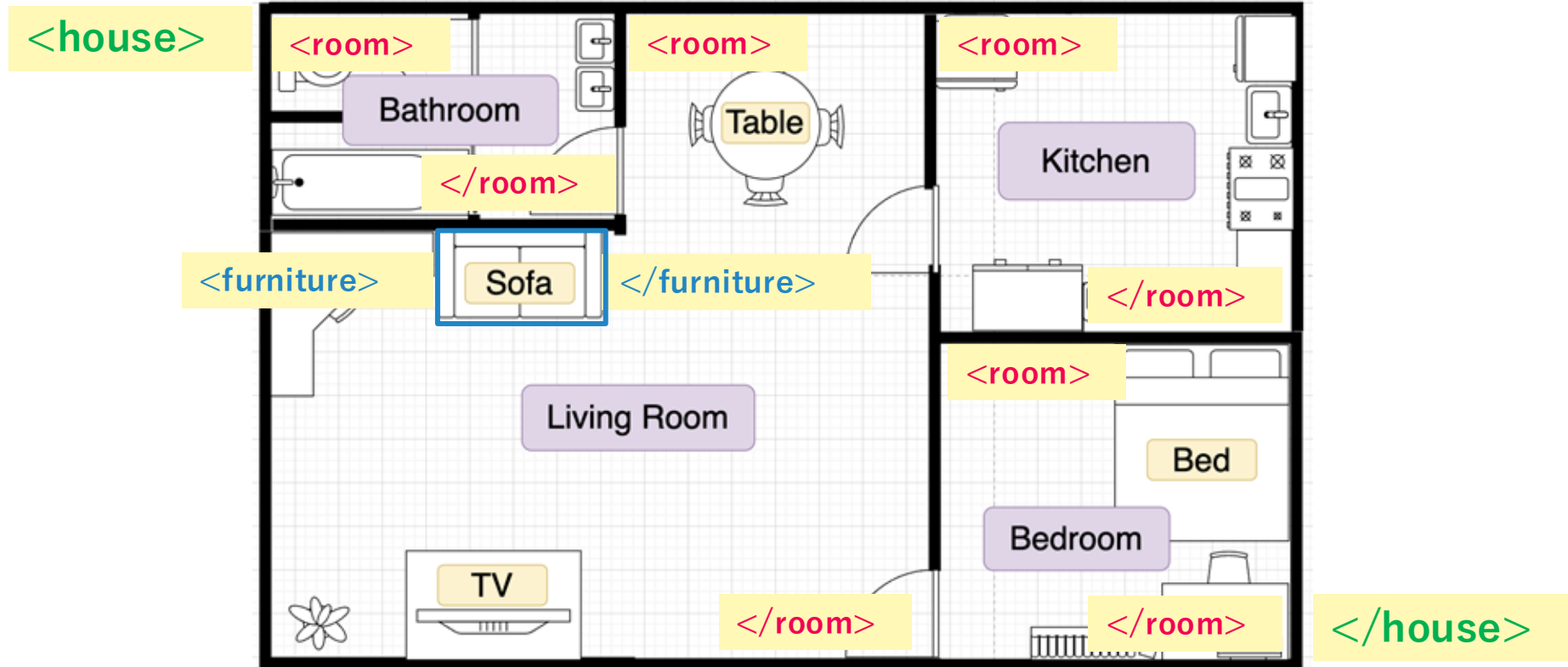
- Bedroom

-- bed

Each room = a section of the text, such as a chapter or paragraph



TEI Structure (2)



TEI Structure (3)

1. Metadata Tags

- Tags that provide **information about the text as a whole** (author, date of writing, place of publication, etc.)

2. Structure Tags

- Tags that indicate **major structural components of the text** (such as chapters, paragraphs, and sections)

3. Content Tags

- Tags that **describe the content of the text** (such as headings, names, dates, quotations, etc.)

TEI Guidelines (1)

TEI Guidelines Homepage

Guidelines for tagging text, its structure, annotations, and other contextual information

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

(Please search for: “TEI guidelines”)

**TEI P5:
Guidelines for Electronic Text
Encoding and Interchange**

by the TEI Consortium
Originally edited by C.M. Sperberg-McQueen and Lou
Burnard for the ACH-ALLC-ACL Text Encoding Initiative
Now entirely revised and expanded under the supervision
of the Technical Council of the TEI Consortium

The TEI Consortium
P5 Version 4.8.0. Last updated on 8th July 2024, revision
1f9891a87

Text Encoding Initiative Consortium
2024

The TEI Consortium

**TEI P5:
Guidelines for Electronic Text Encoding and Interchange**

Text Encoding Initiative Consortium 2024

TEI Guidelines (2)

Goal

The Standardization of Text Data



TEI Guidelines (3)

```
<text>
<body>
  <div type="poem">
    <lg type="stanza">
      <l>In the quiet forest,</l>
      <l>I walk the path lit by moonlight.</l>
      <l>A cold wind blows through,</l>
      <l>and memories stir within me.</l>
    </lg>
  </div>
</body>
</text>
```

↑Poem

Diary→

Letter→

```
<text>
<body>
  <div type="diary">
    <div type="entry" xml:id="entry1">
      <head>September 25, 2024</head>
      <p>Today was an extremely busy day. I was rushed with work from the morning and couldn't even eat lunch.</p>
    </div>
    <div type="entry" xml:id="entry2">
      <head>September 26, 2024</head>
      <p>Today I took things slowly. The weather was nice, and I enjoyed a walk.</p>
    </div>
  </div>
</body>
</text>
```

```
<text>
<body>
  <div type="letter">
    <opener>
      <salute>Dear friend,</salute>
    </opener>
    <p>Sorry I haven't written in so long. Life has been busy lately, but I think of you often.</p>
    <closer>
      <salute>Best regards,</salute>
      <signed>Taro</signed>
    </closer>
  </div>
</body>
</text>
```

TEI Guidelines (4)

TEI Guidelines Homepage →



TEI: Guidelines for Electronic Text Encoding and Interchange

P5 Version 4.8.0. Last updated on 2nd September 2024, revision 4e6e41b0b

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)



Front Matter

[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](#)

Back Matter

- Appendix A [Model Classes](#)
- Appendix B [Attribute Classes](#)
- Appendix C [Elements](#)
- Appendix D [Attributes](#)
- Appendix E [Datatypes and Other Macros](#)
- Appendix F [Bibliography](#)
- Appendix G [Deprecations](#)
- Appendix H [Prefatory Notes](#)
- Appendix I [Colophon](#)

Text Body

- 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](#)
- 15 [Tables, Formulae, Graphics, and Notated Music](#)
- 16 [Language Corpora](#)
- 17 [Linking, Segmentation, and Alignment](#)
- 18 [Simple Analytic Mechanisms](#)
- 19 [Feature Structures](#)
- 20 [Graphs, Networks, and Trees](#)
- 21 [Non-hierarchical Structures](#)
- 22 [Certainty, Precision, and Responsibility](#)
- 23 [Documentation Elements](#)
- 24 [Using the TEI](#)

TEI sourcecode

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