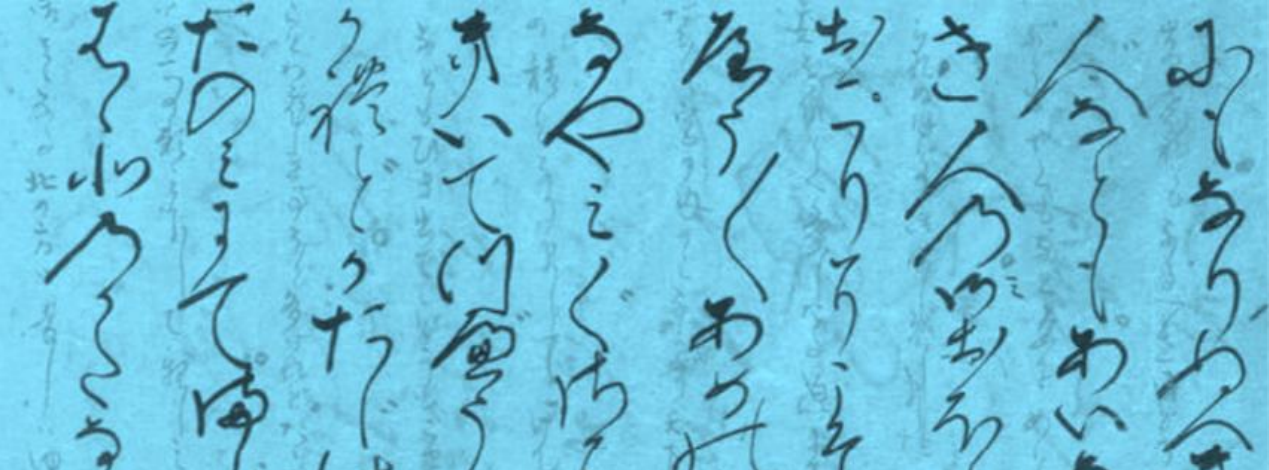


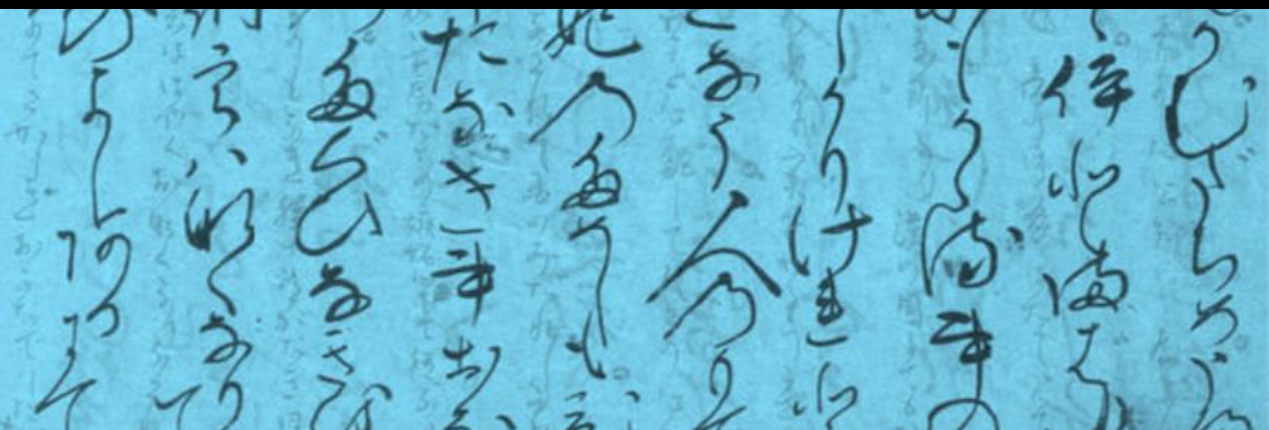


```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>東京大学本源氏物語</title>
7       </titleStmt>
8       <publicationStmt>
```

第8回東京大学学術資産アーカイブ化推進室主催セミナー 2025.1.28

TEI (Text Encoding Initiative) の概要と活用

鈴置 怜子 東京大学情報システム部情報基盤課学術情報チーム デジタル・ライブラリ担当
福嶋 彩乃 東京大学経済学図書館 (図書運用担当)



```
18   </msIdentifier>
19   </msDesc>
20   </sourceDesc>
21   </fileDesc>
22 </teiHeader>
23 <text>
24   <body>
```

自己紹介

◆ 鈴置 怜子

情報システム部情報基盤課学術情報チーム デジタル・ライブラリ担当
担当業務：デジタルアーカイブ・機関リポジトリ

◆ 福嶋 彩乃

経済学図書館 図書運用担当

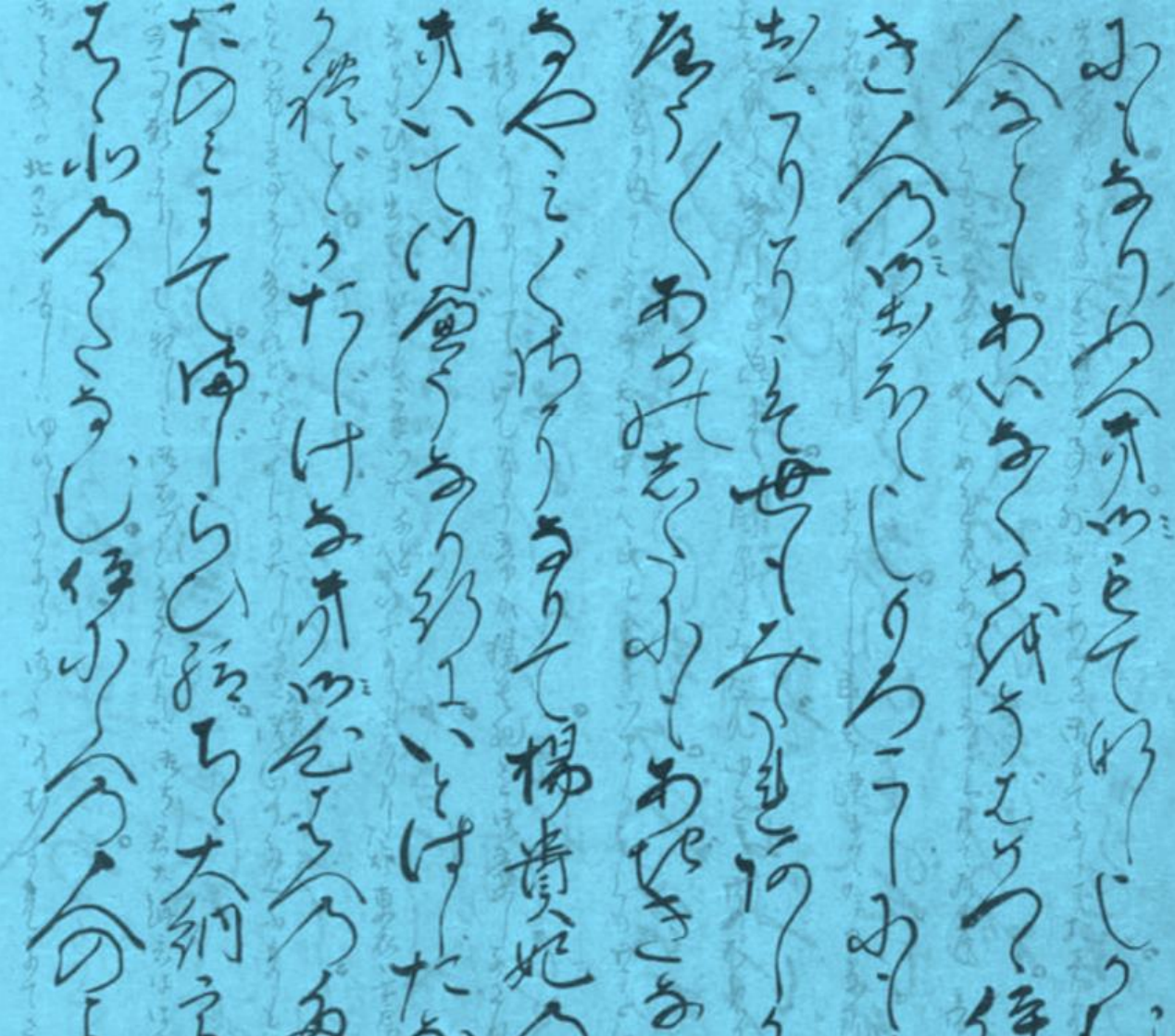
担当業務：サービス業務（カウンター・レファレンス・学外者対応等）

- 東京大学の図書系職員
- 若手図書館員DH勉強会（業務外の自主勉強会）メンバー

<https://dhlibrarianstudygroup.github.io/website/docs/about/>

* DH : Digital Humanities

1. TEIとは何か
2. テキストデータ構造化の利点
3. TEI/XML活用事例
4. TEIマークアップの進め方



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>東京大学本源氏物語</title>
7       </titleStmt>
8       <publicationStmt>
9         <availability>
10          <licence>CC BY 4.0</licence>
11        </availability>
12      </publicationStmt>
13      <sourceDesc>
14        <msDesc>
15          <msIdentifier>
16            <institution>東京大学</institution>
17            <repository>東京大学総合図書館</repository>
18          </msIdentifier>
19        </msDesc>
```

1. TEIとは何か



```
23 <text>
24   <body>
```

TEIとは？

TEI : Text Encoding Initiative



- TEIガイドライン：
人文学のための、電子テキスト構造化の
国際的なガイドライン
- 標準規格ではなく、ガイドラインの提示
- ソフトウェアに依存しないフォーマットを志向

- タグを追記して意味を付与する

- ◆ 囲まれたテキストに意味を付与するタグ

`<title>`平家巻第十一`</title>` title : 「平家巻第十一」がタイトルであることを示す

←→ ←→

開始タグ 終了タグ

- ◆ 挿入した位置に意味を付与するタグ

`<pb/>` page beginning : ページの開始位置であることを示す

←→

空白タグ

XMLの基礎知識

- タグに補足情報を追記することができる

← element (要素) →

<persName corresp="#那須与一">与市</persName>

← attribute (属性) →

personal name : 人物名であることを示す

```
<?xml version="1.0" encoding="UTF-8"?>  
<TEI xmlns="http://www.tei-c.org/ns/1.0">  
  <teiHeader>  
    メタデータ  
  </teiHeader>  
  <text>  
    コンテンツ  
  </text>  
</TEI>
```

XML宣言

TEIの名前空間*

TEIヘッダー

TEIエレメント

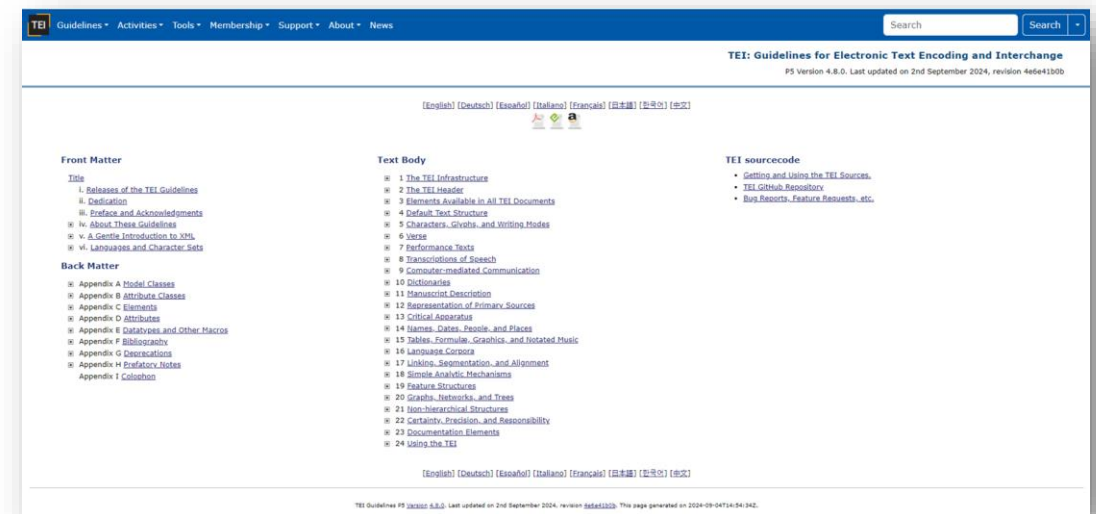
* 名前空間 = 他の用途のタグと名前が衝突しないようにするもの

- スキーマ = 必須のエレメント（要素）などを決めたタグのルール
- スキーマファイルを使って、XML文書の妥当性をチェックする
- スキーマファイルの形式：
 - DTD（拡張子 .dtd）
 - XML Schema（拡張子 .xsd）
 - RELAX NG（拡張子 .rng .rnc）

P5 Version 4.8.0 <https://tei-c.org/release/doc/tei-p5-doc/en/html/>

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, Formulæ, 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ガイドライン 日本語訳

<https://www.dh.ku-orcas.kansai-u.ac.jp/?cat=9>

P5ガイドライン全24章のうち、第1章「TEIインフラストラクチャ」・第2章「TEIヘッダー」・第4章「デフォルトのテキスト構造」の日本語訳（2025年1月現在）

◆ 日本語向けTEIガイドライン ver. 0.2

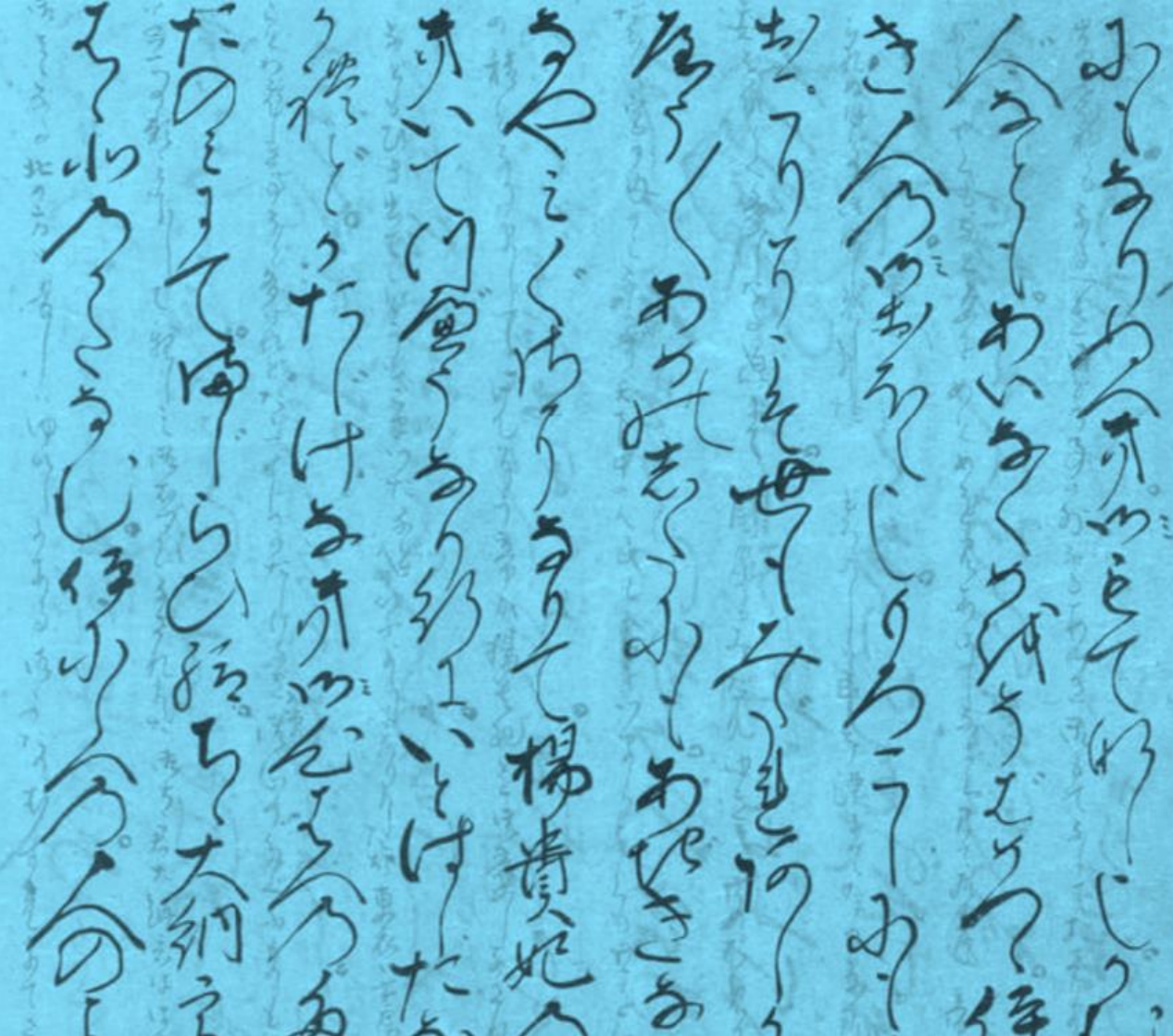
https://github.com/TEI-EAJ/jp_guidelines/wiki

主に日本語の資料をデジタル化するための解説

用途に応じたマークアップ方法やタグの選び方、ツールの紹介

◆ 日本古典籍TEI本文データ作成要領

https://github.com/TEI-EAJ/jpn_classical/blob/master/jpn_classical_guideline.md



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>東京大学本源氏物語</title>
7       </titleStmt>
8       <publicationStmt>
9         <availavility>
10          <licence>CC BY 4.0</licence>
11        </availavility>
12      </publicationStmt>
13      <sourceDesc>
14        <msDesc>
15          <msIdentifier>
16            <institution>東京大学</institut
17            <repository>東京大学総合図書館
18          </msIdentifier>
19        </msDesc>
```

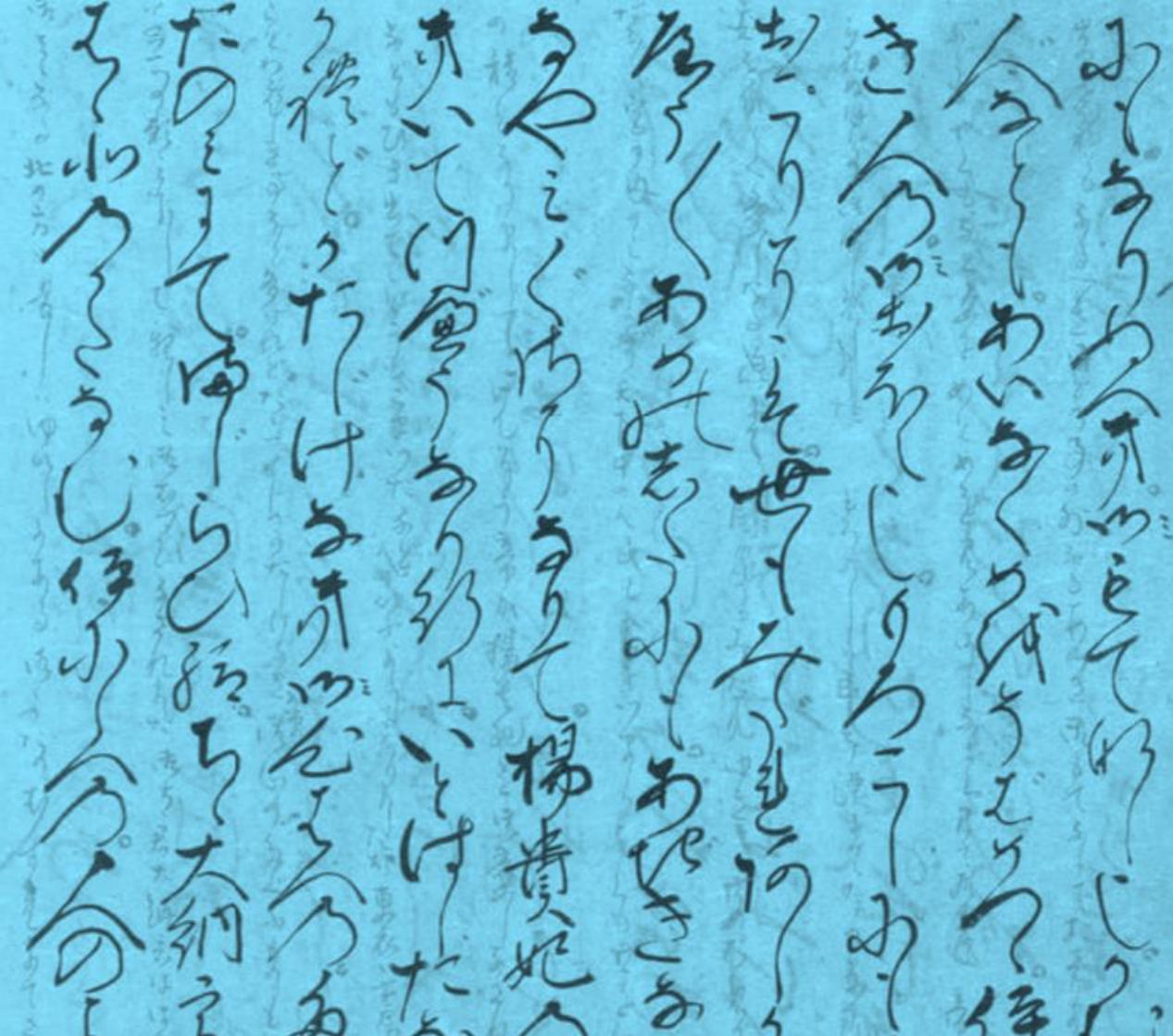
2. テキストデータ構造化の利点



```
23 <text>
24   <body>
```

TEIによるテキストデータ構造化の利点

- 情報の保存
 - 注釈的な情報や解釈を記述できる
 - メタデータとコンテンツをセットにして保存可能
- TEI/XMLの可視化・分析のために開発されたツールを使用できる
 - ビューワ・パラレルコーパス
- 画像データとの連携
 - IIIFマニフェストで対応する画像を記述
- 新たな研究方法の可能性
 - 検索可能性・機械可読性の向上→定量的分析・自然言語処理・大規模言語モデル…
- * 多様な資料のマークアップ事例やTEI/XMLの利活用事例を生み出すことで、TEIのエコシステムを豊かにすることができる



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>東京大学本源氏物語</title>
7       </titleStmt>
8       <publicationStmt>
9         <availability>
10          <licence>CC BY 4.0</licence>
11        </availability>
12      </publicationStmt>
13      <sourceDesc>
14        <msDesc>
15          <msIdentifier>
16            <institution>東京大学</institution>
17            <repository>東京大学総合図書館</repository>
18          </msIdentifier>
19        </msDesc>
```

3. TEI/XML活用事例



```
23 <text>
24   <body>
```

FIHRIST

<https://www.fihrist.org.uk>

- イギリスの大学図書館所蔵イスラーム関連写本のオンライン目録
- 複数の大学が共通のスキームによって目録を作成している
- 書誌情報をTEI/XMLで記述
- 所蔵館、言語、主題、タイトル、作者などで絞り込みでの検索が可能
- TEI/XMLファイルはGitHubにてダウンロード可能
<https://github.com/fihristorg/fihrist-mss>

Union Catalogue of Manuscripts from the Islamicate World


[ADVANCED SEARCH](#)

OR BROWSE BY

[CLASSMARKS](#)
[WORKS](#)
[PEOPLE](#)
[SUBJECTS](#)

Institutions

Oxford University	British Library	Cambridge University	The University of Manchester	School of Oriental and African Studies
Royal Asiatic Society of Great Britain and Ireland	University of Birmingham	King's College Cambridge	Eton College	Wellcome Trust
Wadham College (University of Oxford)	Trinity College Cambridge	University of St Andrews	Arabic Commentaries on the Hippocratic Aphorisms project	Jesus College (Cambridge)

King's Pote 1 (King's College Library, King's College Cambridge)

Pote Collection

Contents

Summary of Contents: 1 copy of *Irshād al-Sālikīn*.

Author: Yūsuf ibn Shaykh Muḥammad Najm al-dīn Gurdīzī یوسف بن شیخ محمد نجم الدین گردیزی

Title: *Irshād al-Sālikīn*

Title: ارشاد السالکین

Language(s): Persian

Colophon details: **Date:** 27 Dhū al-Ḥijjah 1081 / 7 May 1671 **Scribe:** Nūr Muḥammad ibn Ghulām Ḥusayn, known as Qurayshī Multānī نور محمد بن غلام حسین .

Colophon further notes: A note after the colophon is dated 23 Ramaḍān year (11)43.

References

Browne no. 44

Palmer no. 1

Physical Description

Form: codex

Extent: c. 400 folios.

Dimensions (leaf): 28.7 × 17.5 cm.

Dimensions (written): 21.6 × 11.6 cm.

Foliation: Neither foliated nor paginated.

Layout

```
<msItem xml:id="Kings_1-item1" n="1">
  <author key="person_f5738" >
    <persName xml:lang="fa-Latn-x-1c" type="standard">Yūsuf ibn Shaykh Muḥammad Najm al-dīn Gurdīzī</persName>
    <persName xml:lang="fa">یوسف بن شیخ محمد نجم الدین گردیزی</persName>
  </author>
  <title xml:lang="fa-Latn-x-1c" key="work_11031">Irshād al-Sālikīn</title>
  <title xml:lang="fa" key="work_11031">ارشاد السالکین</title>
  <textLang mainLang="fa">Persian</textLang>
  <note>
    <label>Colophon details:</label>
    <label>Date:</label> <date calendar="#Hijri-qamari">27 Dhū al-Ḥijjah 1081</date> / <date calendar="#Gregorian">7 May 1671</date>
    <label>Scribe:</label> <persName xml:lang="fa-Latn-x-1c" key="person_f5739" role="scr">Nūr Muḥammad ibn Ghulām Ḥusayn</persName>
  </note>
  <note><label>Colophon further notes:</label> A note after the colophon is dated 23 Ramaḍān year (11)43.</note>
  <listBibl>
    <bibl>
      <ref target="https://cudl.lib.cam.ac.uk/view/PR-A-00233-00034-00002/13">Browne no. 44</ref>
    </bibl>
    <bibl>
      <ref target="http://menadoc.bibliothek.uni-halle.de/ssg/content/pageview/817964">Palmer no. 1</ref>
    </bibl>
  </listBibl>
</msItem>
```

メタデータがTEI/XMLで書かれていると何が嬉しい？

- 古典籍の書誌情報（料紙、蔵書印、書き入れ、来歴情報などなど）が抽出しやすい（OPAC では備考・注記欄に雑多に記録されることが多い）
- Pythonなどを使って機械的に処理・分析ができる
- Huw, J., and Yasmin, F. (2024) “Manuscript Catalogues as Data for Research: From Provenance to Data Decolonisation”, *Digital Humanities Quarterly*, 18(3).
<https://www.digitalhumanities.org/dhq/vol/18/3/000751/000751.html>
- データ変更の履歴、データ作成者を記録できる = データの信頼性を担保できる

```
<revisionDesc>
  <change when="2020-07-06" xml:id="role-assignment">
    <persName>Julian Cook</persName> Assigned role(s) to every person having an id (Fihrist data cleanup project).
  </change>
  <change when="2018-07-28">
    <persName key="person_f5722" role="cataloguer" change="#role-assignment">Shiva Mihan</persName> created initial record for Fihrist.
  </change>
</revisionDesc>
```

Cambridge Digital Library <https://cudl.lib.cam.ac.uk/>

- ケンブリッジ大学図書館のデジタルライブラリー
- TEI/XMLで記述されたメタデータや本文データを提供
(メタデータ・本文データは1つのTEI/XMLファイルにまとめて提供される)
 - メタデータ：CC BY-NC 3.0
 - 本文データ：CC BY-NC-ND 3.0
- マークアップされたデータをビューワで可視化
- IIIF画像との連携
- データはライブラリアンが作成



Hypoth. 1. Bodies move uniformly
they are retarded by the resistance
ma other force.
Hyp. 2. The alteration of mo
re by wch it is altered.
Hyp. 3.  Motions impress in ^{two} differ

手稿画像 (IIIF)

Home / Browse / Newton Papers / MS Add. 3965

About Contents Thumbnails Transcription Translation More

Transcription

<1r>

Hypoth. 1. Bodies move uniformly in straight lines unless so far as they are retarded by the resistance of y^e Medium or disturbed by some other force.

Hyp. 2. The alteration of motion is ever proportional to y^e force by w^{ch} it is altered.

Hyp. 3.  Two Motions impress in ~~two~~ different lines, if those lines be taken in proportion to the motions & completed into a parallelogram, compose a motion where the diagonal of y^e Parallelogram shall be described in the same time in w^{ch} y^e sides thereof would have been described by the compounding motions apart. The



`<p xml:id="par3">Hyp. 3. <del type="cancelled">Two</del> Motions impress in <add place="supralinear">two</add> different lines,`

TEI/XML

described in the same time in w^{ch} y^e sides thereof would have been described by the compounding motions AB & AC compound the

an immovable center, it shall constantly move in one of the same plane, & in that plane describe equal areas in equal times.

Let A be y^e center towards w^{ch} y^e body is attracted, & suppose y^e attraction acts not continually but by discontinued impressions



Download Image Request Rights Download Metadata IIIF Manifest Open in Mirador

TEI古典籍ビューワ 校異源氏物語冒頭部分

https://candra.dhii.jp/nagasaki/tei/tei_viewer/?file=tales_of_genji_01_01.xml

- 同じタイトルの作品でも、本によって文章が異なることがある（異同がある）
- 従来は複数の文章を並記、あるいは脚注・傍注などで異同を記述してきた

TEI/XMLで記述すると…

- 異同の情報が本単位や文章単位で取り出しやすくなる
- ビューワで表示することで、視認性が向上する

デジタル源氏物語 <https://genji.dl.itc.u-tokyo.ac.jp/>

- 『源氏物語』 本文研究プラットフォーム
- 裏源氏勉強会（東京大学附属図書館職員有志・教員によるデジタルアーカイブ活用に関する勉強会）によるテキストデータ構築・Webサイト公開
- 「画像とテキストを一緒にみる」では、各機関で公開されている複数の諸本画像と『校異源氏物語』・現代語訳（与謝野晶子訳）のテキストを並べて閲覧できる
- ビューワ上で『校異源氏物語』のテキストを選択すると、与謝野晶子訳のテキストにおける対応箇所がハイライトされる

Parallel Text Viewer with TEI & IIIF

校異源氏物語、巻一

東大本

九大本 (古活字版)

九大本 (無跋無刊記整版本)

湖月抄 (国文研所蔵)

Fit Bounds

レイアウト変更

全画面表示

+

アイテム追加

デジタル源氏物語

English

すへたてまつりてをこなふあまりけり。すたれすこしあけて花たてまつる

り。中のはしらにのりてけうそくのうへにきやうをきていとなやましけ

みたるあまきみたゝ人とみえず。四十よはかりにいていしうあてにやせ

たれ

とつらつきふくらにまみのほかのうつくしけにそかれたるす衆も中く

なかりもこよなういまめかしきものとあはれにみ給ふよけな

と

なたりはかりさてはわらはへそいていりあそふ。中に十はかりやあらむと

みえ

てしるきぬ山ふきなどなへたるきてはりきたる女こあたまみえつること

もにへるへうもあらず。いみしくおいさきみてうつくしけなかなかり

かみ

はあふきをひろけたるやうにゆらくとして。かほはいとあかくすりなして

たて

り。なに事そやわらはへとはち給へるかとあまきみのみあけたるにす

こし

おほえたるどころあはれはこめりどみ給ふすめをいぬきかにかしつる

ふせ

このうちにこめたりつるものをどいてくちおしとおもへり。このあたるお

とな

れい心のなしかるわさをしてさなまる。こいといと心つきなけれ。い

かた

へかまかりぬる。いとおかしうやう。なりつるものをからすなともこそみ

つく

れとてたちてゆく。かみゆる。かにいとなくめやすき人なり。少納言の

めと

く。こそいふめるはこのこのうしるみなるへし。あま君いてあなおさなや

ふか

ひなうものし給かな。をのかかくけふあすにおほゆるいのちはなにともお

ほし

たらてすめした給ほとよ。つみうることそとつねにきこゆるをうくく

こ

ちやといへはつあたり。つらつきいとらうたけにまゆのわたりうちけふ

りい

はけなかりやたるひたいつきかむさし。いみしうつくし。ねひゆかむさ

かし

かき人かなとめとまり給。さるはかりなう心をつくしきこゆる人。いに

よう

にたてまつれるかまらるなりけりとおもふにもなみたさおつる。あまきみ

かみ

をかきなてつ。つける事をうさかり給へとおかしの御くしや。いとはかな

うも

のし給。こあはれにうしるめたけれ。かはかりになれはいとくらぬ人もあ

るも

のをこ姫君は十はかりに殿にをくれ給ひしほいみしうものはおもひし給

へり

し。そかし。いまをのれみすてたてまつらはいかて世におはせむとす

むむ

て、その山莊をのそくこの垣根のすぐ前になつてゐる西向きの座敷に持仏を置いてお勤めをする尼がいた。簾を少し上げて、その時に仏前へ花が供えられた。室の中央の柱に近くすわて、脇息の上に経巻を置いて、病苦のあるふうでそれを読む尼はただの尼とは見えな。四十ぐらいて、色は非常に白くて上品に、髪は黒くはいるが、髪はふくりにて、目つきの美しいのとともに、短く切り捨ててある髪、裾のそつたのが、かえつて長い髪よりも艶なものであるという感じを与えた。きれいな中年の女が一人いて、そのほかにこの座敷を出たりはいたりして遊んでいる女の子が幾人かいて、その中に十歳ぐらに見えて、白の上に淡黄の柔らかな着物を重ねて向こうから走って来た子は、さきか何人も見た子供とはいひに言うことのできな麗質を備えていた。将来はどんな美しい人になるだろうと思われるところがあつて、肩の垂れ髪を裾をひろげたようにたくさんでゆらゆらとしていた。顔は泣いたあとのよう、手にすって赤くなつてゐる。尼さんの横へ来て立つ、

「さうしたの、童女たちのことで憤つてゐる。」

こつて見上げた顔と少し似たところがあるので、この人の子なのであろうと源氏は思った。

「省の子を大倉が逃がしてしまいましたの、伏籠の中に置いて逃げないようにはしてあったのに。」

「たいへん残念そうである。そばにいた中年の女が、

「またいつもの粗相やさんがそんなことをしてお嬢様にしかられるのですね、困った人です。省はどちらのほうへ参りました。だいが馴れてきてかわゆくございましたのに、外へ出ては山の鳥に見つかつてごんごんにあわされまうか。」

と言ひながら立つて行つた。髪はゆらゆらと動く後姿も感じのよい女である。少納言の乳母と他の人が言つてゐるから、この美しい子供の世話役なのであつた。

「あなたはまたいつまでも子供供らして困つた方ね。私の命がもう今日明日かと思はれるのに、それは何とも思はないで、省のほうで惜しいのだね。省を籠に入れておいたりすることは仏様の喜びにならないことだと私はいつと言つてゐるのに。」

と尼君は言つて、また、

「こゝろ」と

と言つて美しい子は下へすわつた。顔つきが非常にかわいく、眉のほかに伸びたところ、子供らしく自然に髪が横撫になつてゐる額にも髪は性質にも、すぐれた美がひそんでゐると見えた。大人になつた時を想像してすばらしい佳人の姿も源氏の君は目に描いてみた。なぜこんなに自分の目がこの子に引き寄せられるのか、それは悪い魔の言によつて似てゐるからであると思つた。判然にも、その人への思慕の涙が熱く頬を伝へた。尼君は女の子の髪をなでながら、

「梳かせるのもうさがるけれどよい髪だね。あなたがこんなふうにあまり子供らしいことで私は心配してゐる。あなたの年になればもうこんなふうでない人もあるのに、亡くなつたお姫さんは十二でお父様に別れたのだけれど、もうその時には悲しみも何もよくわかる人になつてゐましたよ。私が死んでしまつたあとであなたはうなるのだらう。」

あまりに泣くので、隙を見てゐる源氏までも悲しくなつた。子供心にもさすがにじつとしばらく尼君の顔をながめ入つて、それからうつむいた。その時に顔からこぼれかゝつた髪がつかつかと美しく見えた。

生ひ立たんあかりかも知らぬ若輩をおくらす露を消えんぞらなき

校異源氏物語

現代語訳 (与謝野晶子訳)

<anchor/>タグを用いて一文単位でのIDを挿入し、対応箇所を関連付けることでハイライト機能を実現

デジタル源氏物語 TEI活用事例④：対応箇所の記述

テキストデータ構築作業

The screenshot displays the TEI Anchor Editor interface. On the left, a text editor shows a list of text segments with corresponding IDs in brackets, such as [YG1600000300]. A red box highlights the text '校異源氏物語テキストデータのGoogleドキュメント'. A blue box highlights the text '挿入された現代語訳の文ID'. On the right, a preview window shows the resulting TEI/XML structure, with a red box highlighting the text '現代語訳のTEI/XMLファイルをCETELceanで表示'. The interface includes a menu bar, a toolbar, and a status bar.

青空文庫で公開されている
与謝野晶子訳のHTMLから
TEI/XMLを作成、文単位で
IDを付与



校異源氏物語テキストデータの
対応箇所にIDを挿入

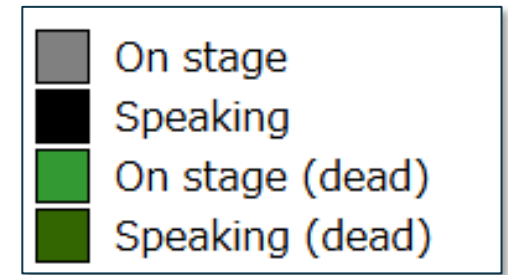
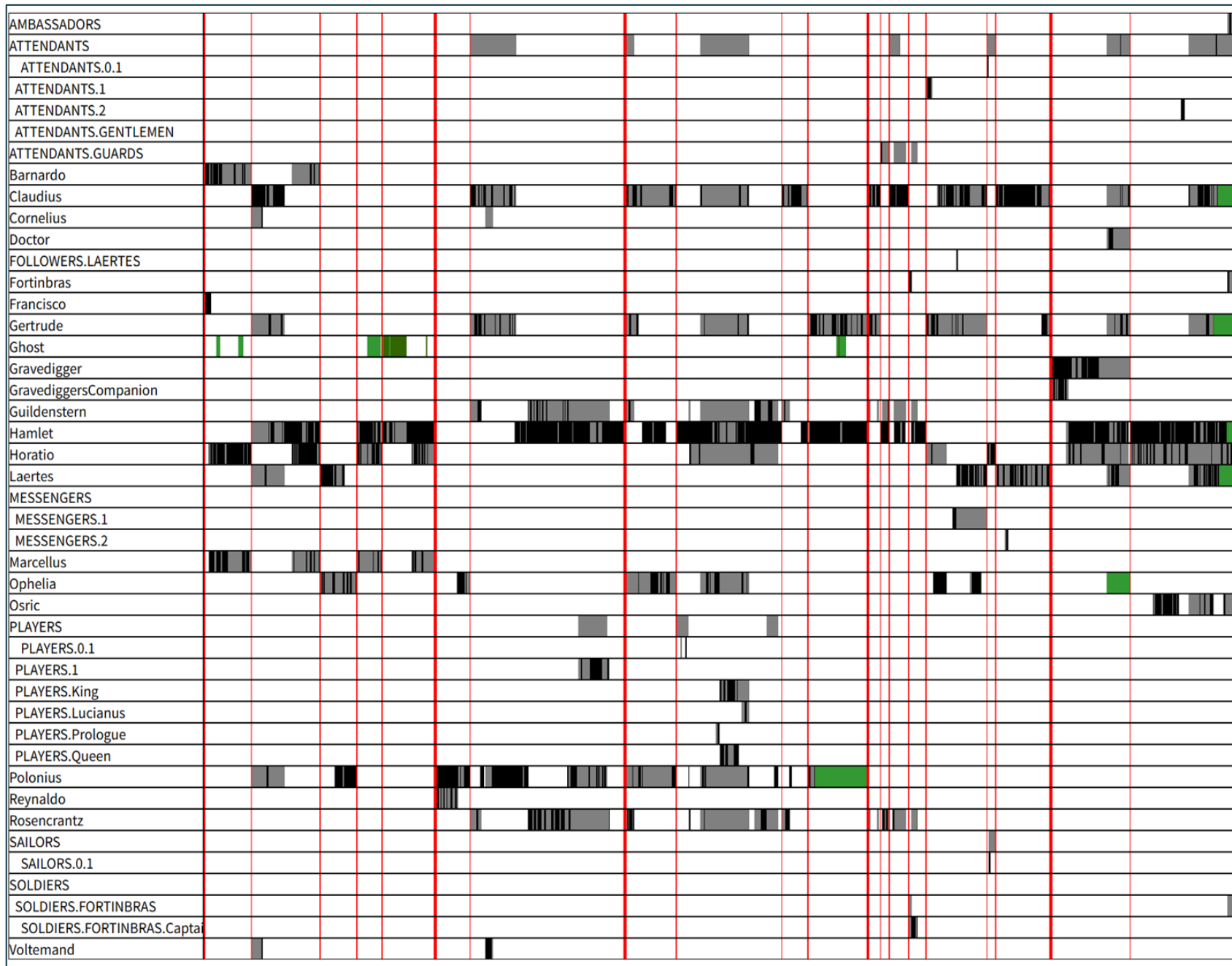


TEI/XMLに変換

中村覚、田村隆、永崎研宣「源氏物語本文研究支援システム「デジタル源氏物語」の開発におけるIIIF・TEIの活用」
『研究報告人文科学とコンピュータ』124巻（情報処理学会、2020年9月）<http://id.nii.ac.jp/1001/00206588/> より引用

The Folger Shakespeare <https://www.folger.edu/>

- フォルジャーシェイクスピア図書館による、シェイクスピア作品のTEI/XML公開
- TEIマークアップによって発話の有無や生死を記述
→各場面における登場人物の状態を抽出・可視化することができる
- テキストデータの提供に加え、可視化の事例を公開している



<https://www.folgerdigitaltexts.org/Ham/charChart>

東京大学のTEI/XML公開事例

東京大学デジタルアーカイブズ構築事業

一部資料のTEI/XMLをGitHubで公開、東京大学デジタルアーカイブズポータルからTEIビューワにリンク

<https://da.dl.itc.u-tokyo.ac.jp/portal/assets/102a847a-ea96-464a-84ba-cb714696bfbf>

The screenshot displays a web-based TEI viewer. On the left, a modern TEI/XML transcription of the document '甲州法度之次第' (Koshu Hoppo no Shiji) is shown in a structured list format. On the right, the original handwritten Japanese document is displayed as a scanned image. The viewer includes navigation controls (plus, minus, download icons) and a status bar at the bottom indicating '1 of 1 • [1]'.

甲州法度之次第

- 國中地頭人等不申子細
悉稱罪科跡私令沒収条
基以自由之至也若犯科
人等為暗信被官者不可有
地頭※田畠之事者加下知可
出別人年貢諸役等地頭江
速可辦償至思地者不及
書載次在家并妻子貢財之
事者如定法職仁可渡之
- 出公事於沙汰場之後奉行
人之外不可致披露況於落着
之儀乎若又未出沙汰場者雖
為奉行人之外不及禁之歟
- 不得内儀而他國へ音物書札遣之
事一向停止之畢但信州在
國人為謀略一國中通用者
無是非次第也若境目之人日
來書狀通來者不能禁歟
- 他國へ結縁嫁或執所領或出
被官契約之条太以為違犯之
基堅堅可禁之若背此旨
輩者可加炳誠也
- 札狼藉田畠事於年貢地者
可為地頭計至思地者以下知可
定之但就賣物等之儀者随分限
可有其沙汰也
- 百姓抑留年貢事罪科不
輕於百姓地者任地頭覺悟可令
所務若有罪科之儀者以檢使
可改之

三 甲州法度之次第

一 國中地頭人等不申子細
悉稱罪科跡私令沒収条
基以自由之至也若犯科
人等為暗信被官者不可有
地頭※田畠之事者加下知可
出別人年貢諸役等地頭江
速可辦償至思地者不及
書載次在家并妻子貢財之
事者如定法職仁可渡之

一 出公事於沙汰場之後奉行
人之外不可致披露況於落着
之儀乎若又未出沙汰場者雖
為奉行人之外不及禁之歟

一 不得内儀而他國へ音物書札遣之
事一向停止之畢但信州在
國人為謀略一國中通用者
無是非次第也若境目之人日
來書狀通來者不能禁歟

一 他國へ結縁嫁或執所領或出
被官契約之条太以為違犯之
基堅堅可禁之若背此旨
輩者可加炳誠也

一 札狼藉田畠事於年貢地者
可為地頭計至思地者以下知可
定之但就賣物等之儀者随分限
可有其沙汰也

一 百姓抑留年貢事罪科不
輕於百姓地者任地頭覺悟可令
所務若有罪科之儀者以檢使
可改之

1 of 1 • [1]

東京大学法学部研究室図書室法制史資料室所蔵『甲州法度之次第』 <https://utda.github.io/tei-viewer/?v=true&u=https%3A//archdataset.dl.itc.u-tokyo.ac.jp/collections/law/text/102a847a-ea96-464a-84ba-cb714696bfbf/102a847a-ea96-464a-84ba-cb714696bfbf.xml>

東京大学のTEI/XML公開事例

ゲーテ自署付書簡テキストデータ

<https://utda.github.io/goethe/>

書誌情報

タイトル

日本語

東京大学総合図書館所蔵のゲーテ自署付書簡 [Letter] Weimar, 29. Dcbr. 1822 [to L. W. Cramer] (WA IV-36, S. 249対応)

ドイツ語

Der Brief von Goethe an Ludwig Wilhelm Cramer vom 29. Dezember 1822 im Besitz der Universitätsbibliothek Tokio

著者

ドイツ語

Johann Wolfgang von Goethe [VIAF](#)

日本語

ヨーハン・ヴォルフガング・フォン・ゲーテ

共著及び清書

ドイツ語

Johann August Friedrich John

日本語

ヨーン

文献記号について

WA

日本語

ヴァイマル版ゲーテ全集 (略称WA)

ドイツ語

Goethes Werke. herausgegeben im Auftrage der Großherzogin Sophie von Sachsen

UTL

日本語

東京大学総合図書館所蔵のゲーテ自署付書簡

ドイツ語

Der Brief von Goethe an Ludwig Wilhelm Cramer vom 29. Dezember

東京大学総合図書館所蔵のゲーテ自署付書簡

Ew: Wohlgeb. .

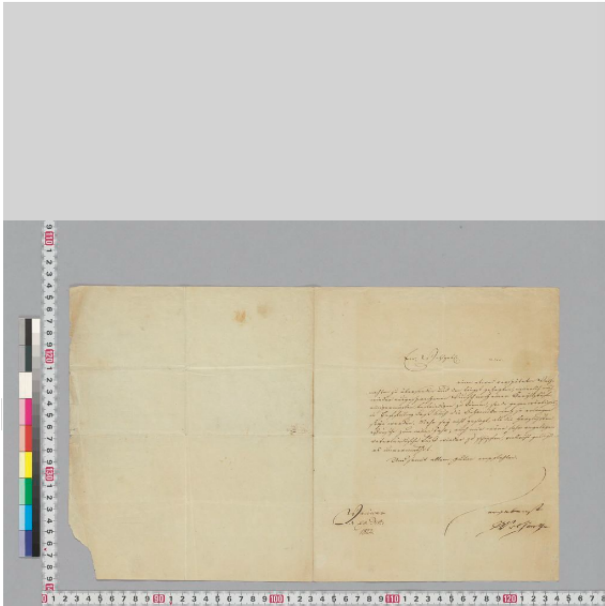
einen etwas verspäteten Weih=
nachten zu übersenden und den längst gehegten, neuerlich auch
wieder ausgesprochenen Wunsch nach einer Beryllstufe
einigermaßen befriedigen zu können, sende **gegenwärtiges**
in Hoffnung daß auch die **Bohemika** noch zu erlangen
seyn werden. Mehr sey nicht gesagt, als die herzlichsten
Wünsche zum neuen Jahr; auch mir wäre sehr angelegen
vaterländische Luft wieder zu schöpfen, vielleicht gelingt
es unvermuthet.

Und somit allem Guten empfohlen.
Weimar
d. 29. Dcbr. 1822.

ergebenst

JW v Goethe

画像



その他の事例

◆ 蔣洲咨文

<https://www.hi.u-tokyo.ac.jp/collection/digitalgallery/wakozukan/tei/>

- ・ 東大史料編纂所蔵の『蔣洲咨文』の画像、釈文、読み下し、大意のテキスト、地図、人物・地名のリストを並行して閲覧できる
- ・ 右上の「ダウンロード」からTEI/XMLファイルを確認できる

◆ 策彦周良文集

<https://www.hi.u-tokyo.ac.jp/collection/digitalgallery/sakugenbunshu/>

- ・ 東大史料編纂所蔵の『策彦周良文集』の画像、翻刻文
- ・ ビューワ上の固有名詞や返り点などの表示／非表示を選択できる

◆ SAT大蔵経2018

<https://21dzk.l.u-tokyo.ac.jp/SAT2018/master30.php>

◆ 十番虫合絵巻ビューワ：https://juban-mushi-awase.dhii.jp/tei_viewer/

◆ 廣瀬本万葉集TEI/XMLデータ分析入門：<https://dhportal.ac.jp/?p=1533>

- ・ Pythonを用いて万葉集のTEI/XMLデータ解析が実践できる（Google Colab上で実行）

その他の事例

- ◆ 渋沢栄一ダイアリー：<https://shibusawa-dlab.github.io/app1/>
 - ・『渋沢栄一伝記資料』に出てくる人物名や地名、日時などを可視化
- ◆ 論語と算盤オンライン：<https://eiichi.shibusawa.or.jp/features/rongotosoroban/index.html>
 - ・渋沢栄一『論語と算盤』の全文をテキスト／画像で閲覧できる
 - ・TEI/XMLファイルは近日公開予定
- ◆ British National Corpus：<http://www.natcorp.ox.ac.uk/>
 - ・品詞情報を付与したイギリス英語のコーパス
- ◆ Transcribe Bentham：<https://blogs.ucl.ac.uk/transcribe-bentham/>
 - ・ユニバーシティ・カレッジ・ロンドン（UCL）が所蔵するJeremy Benthamの遺稿をTEIに準拠する形で翻刻
 - ・翻刻はボランティアによって行われる
 - ・このテキストを基にした全集が刊行中
- ◆ Digital Library for Dutch Literature (DBNL)：<https://www.kb.nl/en/research-find/datasets/dbnl-dataset>
 - ・オランダの文学、言語学、文化史のテキスト
 - ・テキストは著者、タイトル、地名等の関連情報への参照を含むTEI/XML形式

- メタデータの記述と流通

本文情報と書誌情報をひとつのXMLファイルとして流通させることが可能

- 公開データの利活用促進

構造化によって機械可読性の高いデータを提供できる

- デジタルアーカイブ

IIIFマニフェストをTEI/XMLに埋め込むなど、IIIF画像との親和性が高い

- 学問分野を超えた教育の場

欧米圏では、ライブラリアンがDigital Humanitiesの教育を担当している事例がある

- 情報基盤の提供

研究者個人では維持・管理が難しいデータ公開基盤

- 研究データ管理

オープンサイエンスの推進と大学図書館



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>東京大学本源氏物語</title>
7       </titleStmt>
8       <publicationStmt>
9         <availavility>
10          <licence>CC BY 4.0</licence>
11        </availavility>
12      </publicationStmt>
13      <sourceDesc>
14        <msDesc>
15          <msIdentifier>
16            <institution>東京大学</institut
17            <repository>東京大学総合図書館
18          </msIdentifier>
19        </msDesc>
```

4. TEIマークアップの進め方



```
23 <text>
24   <body>
```

TEIマークアップ作業の進め方

■ エディタを使ってタグ付け

- Oxygen XML Editor



有料。TEIのスキーマが組み込まれており、編集時のXML文書がスキーマに沿っているか検証できる。

- Visual Studio Code



無料。XML編集に特化したエディタではないが、拡張機能で入力をサポートできる。

■ タグを意識しない形で作業→XMLに変換

- プログラム処理

タグへの変換やXML文書への変換を一括処理する。

* タグ付けの対象とする情報や作業者の環境次第で、様々な進め方が考えられる。

Best Practices for TEI in Libraries

<https://web.archive.org/web/20230102143254/https://tei-c.org/extra/teiinlibraries/4.0.0/bptl-driver.html>

(※ Wayback Machine にて閲覧)

- TEI協会の図書館分科会によるルール（分科会は現在休会中）
 - タグ付けのレベルを Level 1 - Level 5 の5段階に整理
 - MARCとTEIヘッダーの対応表、それぞれのレベルで求められるタグの具体例も提示
-
- Level 1 : OCRによるテキスト生成と自動化できる範囲でのタグ付け
 - TEIヘッダー（メタデータ）がテキストファイルに添付されることが主な利点
 - レベル2以上のタグ付けの基礎となる
 - 電子テキストのみで提供することは意図されていない（ページ画像とともに提供される）
-
- Level 2 : 最小限のテキストの構造をタグ付け
 - 検索を向上させるための簡単な階層構造を特定するためのタグ付け（テキストの区切りや見出し等）
 - OCRの読み取り結果を校正することは期待しない
 - 原則として電子テキストのみでは提供されない

Best Practices for TEI in Libraries

■ Level 3：簡単な整理を含むタグ付け

- ・人による編集が必要とされるが、テキストの論理構造と体裁に関するタグ付け（ページの表裏、テキストの区切り、段落区切り、図、表など）を目的とするため、専門的な内容分析は行われない
- ・電子テキストのみでも独立可能

■ Level 4：内容に関する基本的な分析を含むタグ付け

- ・テキストの外観（段落など）だけでなく、内容を記述するタグ（人名・地理情報・日付情報など）も付与される

■ Level 5：学術的プロジェクトのためのタグ付け

- ・研究に活用するための、深く分析されたテキストを作成

<http://dhlibrarianstudygroup.github.io/website/docs/about/>



■ DHに興味を持つ若手図書館員による私的な勉強会

- 鈴置 怜子 東京大学情報システム部 情報基盤課学術情報チーム デジタル・ライブラリ担当
- 西口 真梨奈 東京都立中央図書館 サービス部 情報サービス課 サービス計画担当
- 福島 彩乃 東京大学経済学図書館 図書運用担当
- 星合 耀子 国文学研究資料館 管理部 学術情報課 データ標準化推進係
- 村田 祐菜 国立国会図書館 電子情報部
- 吉村 雄太 国立国会図書館 利用者サービス部

■ 読書会（2023年～）

- Arianne, H., Laura, B and Liorah, G (Eds.): Digital Humanities in the Library: Challenges and Opportunities for Subject Specialists, Association of Research Libraries (2015).
- DHに関する論文を個々人で持ち寄り

■ TEIマークアップの実践（2024年～）

- VSCodeでマークアップするための環境設定
- GitHubでの共同編集

① 平家物語 百二十句本「扇の的」

- 京都府立京都学・歴彩館蔵：<https://kokusho.nijl.ac.jp/biblio/100171975/704>
- 国立国会図書館蔵：<https://dl.ndl.go.jp/pid/2546105/1/15>

- 古典籍の書誌情報（メタデータ）
- 異同・人物名・地名のタグ付け

② 抽齋年譜

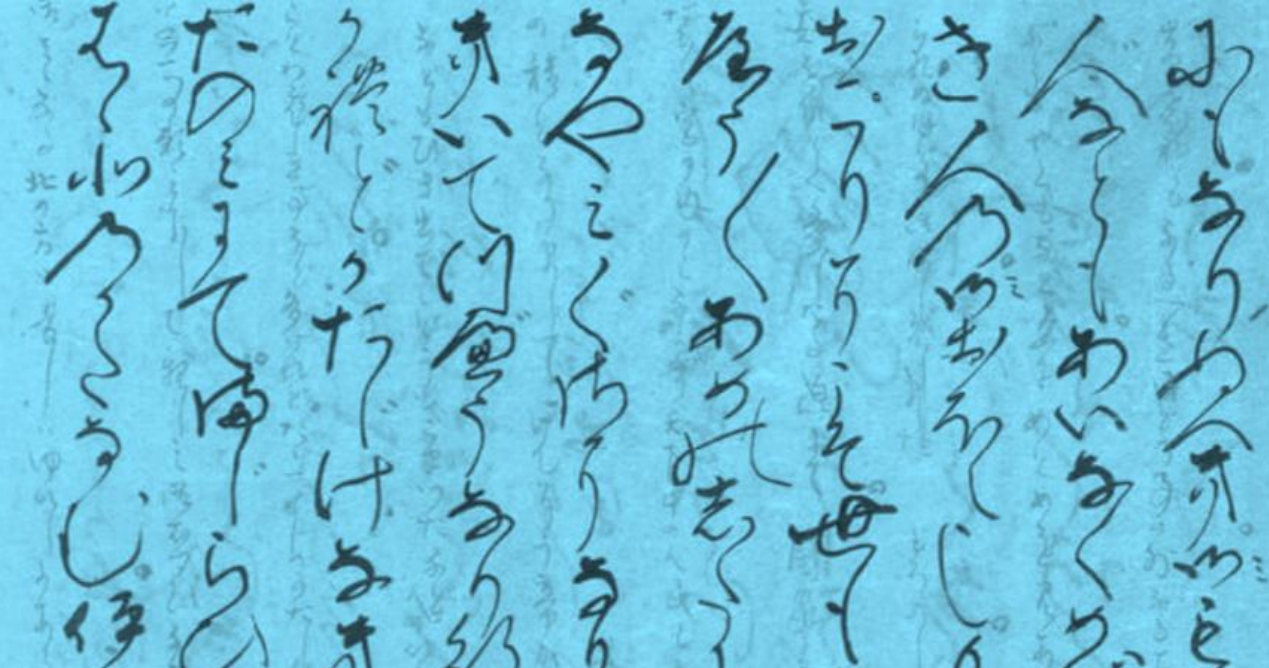
- 東京大学総合図書館蔵：
<https://da.dl.itc.u-tokyo.ac.jp/portal/assets/60d193c1-d814-42c3-90c0-a5caf5975e1a>

参考文献

- ◆ 一般財団法人人文情報学研究所監修、石田友梨・大向一輝・小風綾乃・永崎研宣・宮川創・渡邊要一郎 編『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』文学通信、2022年
- ◆ 裏源氏勉強会 (2024) 「TEIで諸本を繋ぐ：「デジタル源氏物語」の試み」
<https://repository.dl.itc.u-Tokyo.ac.jp/records/2010262>
- ◆ ケンブリッジ大学デジタル図書館がやって来る！ <https://u-parl.lib.u-Tokyo.ac.jp/Japanese/cambridgelibrary20230218>
- ◆ 須永恵美子「イスラーム写本横断検索サイトFIHRISTのご紹介」 <https://u-parl.lib.u-Tokyo.ac.jp/Japanese/column68>
- ◆ 永崎研宣「TEIガイドラインとは何か」 <https://doi.org/10.51094/jxiv.22>
- ◆ 永崎研宣「ケンブリッジ大学デジタル図書館の日本資料の書誌情報を視覚化してみる」
<https://digitalnagasaki.hatenablog.com/entry/2021/02/20/132545>
- ◆ 永崎研宣 (2019) 「歴史データのさまざまな応用 – Text Encoding Initiative の現在 –」『歴史情報学の教科書：歴史のデータが世界をひらく』 <https://bungaku-report.com/blog/2019/03/chapter-8-text-encoding-initiative.html>
- ◆ 永崎研宣、幾浦裕之、藤原静香「古典本文をWEBに載せる」『古典の再生』、文学通信、2024年

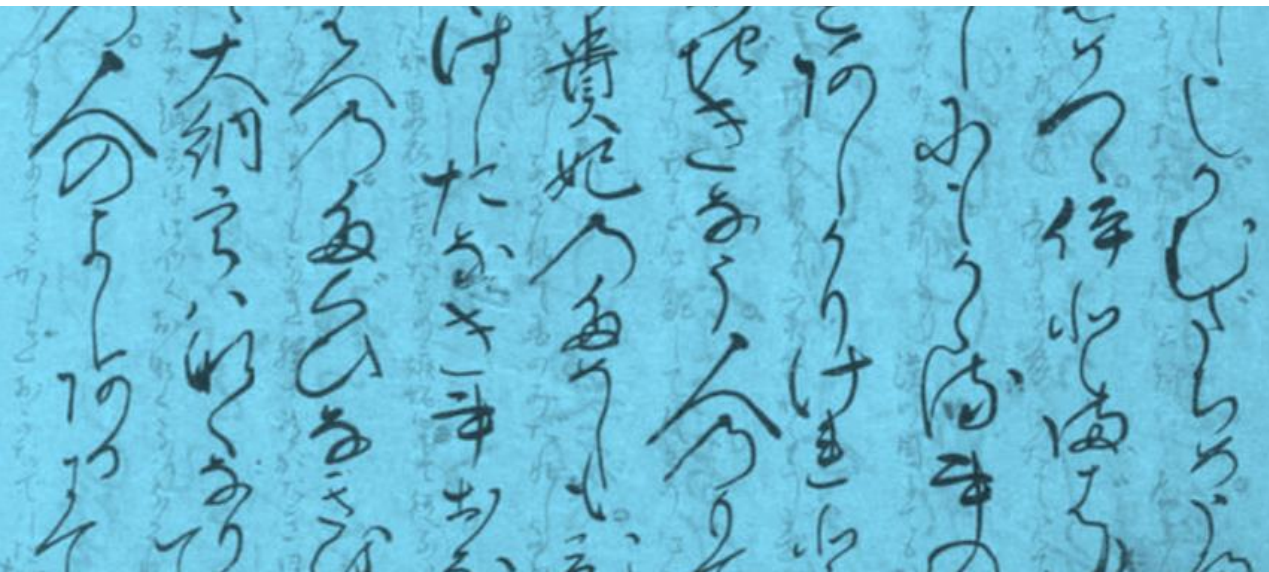
参考文献

- ◆ 中村覚、田村隆、永崎研宣 (2020) 「源氏物語本文研究支援システム「デジタル源氏物語」の開発におけるIIIF・TEIの活用」『研究報告人文科学とコンピュータ』124巻 <http://id.nii.ac.jp/1001/00206588/>
- ◆ 中村美里、木越みち、小川夏代子 (2024) 「研究プラットフォーム構築のためのデジタルデータ活用—教職協働による「デジタル源氏物語」公開の試み—」『大学図書館研究』126巻 <https://doi.org/10.20722/jcul.2165>
- ◆ DARIAH (欧州人文学デジタル研究基盤) MOOC “Digital Scholarly Editions: Manuscripts, Texts and TEI Encoding” <https://www.youtube.com/playlist?list=PL77mHK9JuenN9NXeXQbVcUORz7HZk-9Pv>
- ◆ Ide, Nancy M., C. M. Sperberg-McQueen, Lou Burnard (2018) 「TEI：それはどこからきたのか。そして、なぜ、今もなおここにあるのか？」『デジタル・ヒューマニティーズ』1巻 https://doi.org/10.24576/jadh.1.0_3
- ◆ TEI-EAJ 「IIIF画像とのリンク」 https://github.com/TEI-EAJ/jp_guidelines/wiki/IIIF%E7%94%BB%E5%83%8F%E3%81%A8%E3%81%AE%E3%83%AA%E3%83%B3%E3%82%AF



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>東京大学本源氏物語</title>
7       </titleStmt>
8       <publicationStmt>
9         <availability>
10          <licence>CC BY 4.0</licence>
11        </availability>
```

ご清聴ありがとうございました



```
15   </msIdentifier>
16     <institution>東京大学</institution>
17     <repository>東京大学総合図書館</repository>
18   </msIdentifier>
19   </msDesc>
20 </sourceDesc>
21 </fileDesc>
22 </teiHeader>
23 <text>
24   <body>
```