『延喜式』へのTEI適用と日本史資料のテクストデータ共有・流通

Application of TEI to *Engishiki* and Japanese History Text Data Sharing

小風尚樹・後藤　真

KOKAZE Naoki and GOTO Makoto

［論文要旨］

本論文では，『延喜式』の本文情報のデジタル化と流通の手法について検討を行った。とりわけ TEI（Text Encoding Initiative）という，国際標準を適用し作成したデータについての説明を行い，さらにより広く日本の歴史資料のテクストデータ共有のありようについても述べた。

なお，本研究についての具体的な内容については，すでにいくつかの国際会議等でも発表を行うとともに，論文化も予定されている。そのため，本論文では，これらの技術的側面には詳細に触れることなく，より歴史学の立場からの意義について検討を行った。

筆者らはテクストデータの国際的流通と研究での高度活用を目指し，延喜式の TEI マークアップを行うこととした。TEI（Text Encoding Initiative）とは，人文学に関するテクスト資料を国際的に流通・共有・活用することを目指したプロジェクトであり，そこで作られた規格のことも呼称する。TEI は人文情報学研究の一つの手法として作られているため，歴史資料をどのように理解し，データを加えたかなどの情報をエレメント（タグ付の要素）によって記録することができる点が大きな利点である。このようなメリットに鑑み，筆者らは TEI によるデータ化を提案した。

特に『延喜式』の量的なデータについて TEI によるマークアップをほどこし，トランザクショノグラフィの手法を用い，全体像を解析する可能性について，踏み込んだ検討を行ったほか，合わせてこれらのマークアップ手法を基盤データとして用いるためのマニュアルの作り方について検討を加えた。

日本史研究の活性化という観点からは，このような歴史資料や研究手法の可視化は欠かすことができない。人文学や歴史学が「危機」と呼ばれる現在であるからこそ，基盤データを構築し，自由に流通し，様々な可能性を開く研究を検討することが求められる。

【キーワード】　TEI，延喜式のマークアップ，情報基盤構築，歴史情報学，人文情報学

はじめに

本論文では，『延喜式』の本文情報のデジタル化と流通の手法について，検討を行う。とりわけ TEI（Text Encoding Initiative）という，国際標準を適用し作成したデータについての説明を行うが，それのみならず，より広く日本の歴史資料のテクストデータ共有のありようについても述べるものである。

本デジタル化プロジェクトの全体像を説明し，次に TEI の具体的な中身について述べる。そして，後藤が改めて本プロジェクトに関連するデータ構築手法について述べ，最後に本デジタル化の意義と課題について説明を行うものとする。

なお，本研究についての具体的な内容については，すでにいくつかの国際会議等でも発表を行うとともに，論文化も予定されている。そのため，本論文では，これらの技術的側面には詳細に触れることなく，より歴史学の立場からの意義について検討を行うものである。

1 日本史資料テクストデータの現状

日本史資料のデジタルデータ化，特に資料目録については東京大学史料編纂所をはじめとして，資料所蔵機関のデータはある程度揃っている[(1)]。また，画像データについては，東寺百合文書などの高精細画像[(2)]をはじめ，近代資料では国立公文書館の画像データ[(3)]などもあり，歴博からも中世文書の画像データが公開されるなどの状況がある。また，「日本語の歴史的典籍の国際共同研究ネットワーク構築計画（略称：歴史的典籍ＮＷ事業）[(4)]」においては日本の典籍画像データの公開が広く進められている。

一方，テクストデータは，その作業の煩雑さという観点や，いわゆる「翻刻」の手間と正確性の課題などから，決して多くのデータが流通しているとは言い難い状況である。その中でいくつかの先行事例を確認したい。東京大学史料編纂所は，『大日本史料』などのテクストデータをコンコーダンスとして提供している点は注目に値する。また，国立公文書館の画像データには冒頭の 300 文字について「検索用」として公開されているテクストデータがある。さらに，これらの状況を打開すべく動いているものとしては近世の地震資料の翻刻プロジェクトである「みんなで翻刻[(5)]」などのクラウドソーシング手法があり，「みんなで翻刻」はすでに 500 万文字を超えるテクストデータが集積されている。

しかし，これらのデータについては「みんなで翻刻」を除き，全文が流通しうるものは，ほとんど存在しない。また「みんなで翻刻」もあくまでもクラウドにおいてテクストを構築するものであり，TEI 等については，今後の検討が必要なものである。その点において，日本史学のまとまったテクストの提供はいまだに多いとはいえない。さらにいえば，国際標準にのっとったテクストデータの流通等はなされていないといえる。日本史学に限らなければ，例えば大正新脩大蔵経データベース（SAT）のような事例が存在する[(6)]。しかし，現時点では多分野の例も決して多いものではない。

そのような状況の中，筆者らはテクストデータの国際的流通と研究での高度活用を目指し，延喜式の TEI マークアップを行うこととした。

2 TEI とその意義

まずは，TEI とは何か，という点から述べておきたい。TEI（Text Encoding Initiative）とは，人文学に関するテクスト資料を国際的に流通・共有・活用することを目指したプロジェクトであり，そこで作られた規格のことも呼称する(7)。本稿では，後者の意味で用いることとする。この規格については，ガイドラインが公表されており，現在は P5 と呼ばれるガイドラインが最新版として用いられている。この TEI を用いる理由は，以下のとおりである。

1. テクストが国際標準となるマークアップで作られるため，データの流通が容易。日本語のテクストがある場合，基本的には，最低限の情報を取得する場合においてさえそのテクストの意味を理解しなければならないが，例えば「この部分は量を示している」などの指示がコンピュータによってマークアップされていれば，テクストが詳しく読めなくても，最低限の意味情報を取得できる。
2. 同じく国際標準で作られるため，活用する際に専用のソフトウェアを用いなくてもよい場合がある。同じルールで作られたテクストであれば，同じソフトウェアでの分析が可能となりうる。そのため，より容易にテクスト解析を行うことができる可能性が高まる。さらに，表示方法なども，既存のものを活用しうる。
3. データコンバートが容易になるため，長期的なデータ保存への可能性をひらくことができる。データベース等で公開されているデータ類の長期的な運用が課題となっている。その際には，データとシステムを可能な限り分離してつくることが求められる。また，データの構造はシステムに依存しないものとして作ることが求められる。一般的にテクストデータそのものは，シンプルな構造であり，長期保存が可能なものである。しかし，データベース等のシステムに入れる際には何らかの処置を施す必要が生じる。その際に国際的な標準に従い作っておくことで，システムに依存せずにデータを構築することができる。このことにより，システムの大きな変更があっても，より容易にデータだけは移行することができる。また，最悪の場合として，システムが廃棄されるような事態になったとしても，データを抽出することによりデータを救出することができる。それにより，資料に関するデータは維持されることになる。

これらの利点を持っている点に加え，TEI は人文情報学研究の一つの手法として作られているため，歴史資料をどのように理解し，データを加えたかなどの情報をエレメント（タグ付の要素）によって記録することができる点が大きな利点である。例えば，

```
<head ana=" 食法 "/>
    <measure type=" アワビ " commodity=" 東鰒 "> 東鰒
<num value="2"/><unit ref="# 両 "/></measure>
```

と記述することで，東鯷をどのように認識し，処理をしようとしたかを理解することができる。この場合，斎宮式の食法において，鯷の物品を分類し，記述してある単位（この場合は「両」）で，数的な処置を施そうとしたことがわかる。この場合は，延喜式の分析としては基本的な操作であると理解できるが，そのような操作を行った，という研究の途中作業の記録にもなりうる。そして，この記録を再度利用して，別の研究へと活用することが可能になるのである。このように，研究史を論文以外に個別の手法として残しておける点も大きなメリットである。無論，このような記録は独自のタグやエレメントセットでも可能ではあるが，独自のタグの場合，エレメントがどのような意味を持っており，どのような構造を持っているのかを，再解釈する必要が生じる。一方で，国際標準であれば，そのような再解釈が必要ない点がメリットである。そして，この国際標準で記述しきれないものについては，独自のものを追加できるため，標準と独自のものとを並列させられる点も大きなメリットであるといえよう。

このようなメリットに鑑み，筆者らはTEIによるデータ化を提案した。このプロジェクトは，著者の一人である小風を中心とし，歴博における延喜式の検討チーム，総合資料学の人文情報に関係するチーム，そしてTEIに造詣の深い永崎研宣氏の助言をいただき，構築をしたものである。

1. 基礎的な構造について，一部の式を例にマークアップをほどこす
2. 機械的に一括でできる部分については，全体に機械的に処理を行う
3. 特に小風の研究に深く関わる部分については，小風の研究に即したマークアップを実験的に行い，モデル事例を積み上げる

このような順番によって，データ作成を行った。次章でどのような作業を行ったかを述べることとする。

❶……プロジェクトの具体像

本章では，TEIでマークアップするプロジェクトの作業過程について具体的に紹介する。そこで，まず延喜式研究は門外漢である小風がどのようにしてプロジェクトに関わるようになったのか，次に延喜式がどのような歴史資料なのか，そしてなぜTEIに準拠しなければならないのか，という点について確認しておきたい。

1.1 技術協力者として携わるようになった経緯

小風は，2016年9月から本プロジェクトに技術協力者として携わっている。小風の専門は19世紀イギリス海軍の歴史を外交的側面から考察することであるが，2014年から東京大学大学院横断型教育プログラム「デジタル・ヒューマニティーズ（DH）」を副専攻として修めてきた。特に，人文学史資料のテクストをコンピュータ可読形式で構造化するための国際的枠組みTEI（Text Encoding Initiative）を専門とし，中でも財務記録史料を構造化する方法論「トランザクショノグラフィ（Transactionography）」に関する研究成果を発信してきた[8]。結論を先取りすれば，このトラ

ンザクショノグラフィが，延喜式のテクストデータベースを構築するにあたって重要な役割を果たすのである。

そもそも，財務記録史料というのは，広義の商取引に関する情報を残す歴史資料群のことで，複式簿記や帳簿，領収書や日記から，貿易統計など非常に多岐にわたるものである。トランザクショノグラフィは，散文形式だけでなく複式簿記のような複雑な構造を含む財務記録マークアップのための拡張 TEI スキーマの開発を含む研究であり，その本質は，財務取引を「モノの移動」として構造的に捉えることにある。研究プロジェクトの運用面については，アメリカとドイツの大規模な研究助成を受けており，研究コミュニティ MEDEA（Modeling semantically Enriched Digital Edition of Accounts）によるワークショップが数度開催されてきた。2016 年 4 月にアメリカで開催されたワークショップの研究発表例としては，中世フランスの王侯貴族による森林地帯の管理・運営に関する手稿帳簿，ドイツ騎士団における財務文書，アメリカのプランテーション産業における奴隷貿易管理のための帳簿など，欧米圏の歴史資料を扱ったプロジェクトが多かったが，小風は 19 世紀後半のイギリスと清朝中国の間で行われた軍艦売買のやり取りに関する財務記録史料を扱った[(9)]。このアメリカでの発表を発展させる形で行った日本国内での研究報告がきっかけとなり[(10)]，小風が延喜式のテクストデータベース構築プロジェクトに関わるようになった。

1.2 延喜式とトランザクショノグラフィ

本節では，トランザクショノグラフィと延喜式の関係性について，より詳しく説明していきたい。延喜式は言うまでもなく，10 世紀前後の律令制下の日本における「行政マニュアル」であり，非常に広範な社会的側面に関わる細則が定められた，全 50 巻の編纂史料である。細則の例としては，日本各地の祭式儀礼やそこで必要とされた供物の指定，律令政府に収めるよう定められた租庸調や貢納品の詳細，そして各国に運用資金として割り当てられた正税や公廨稲の額の規定などが挙げられる[(11)]。

このように，延喜式が「行政マニュアル」という性質を持つ以上，特に主計式や主税式に典型的に見られるように，各地の特産品に基づく貢納品の規定や租税の徴収など，必然的に「モノの移動」として理解できる記述が豊富に含まれている。すなわち，延喜式のテクストデータベースを構築するにあたっては，小風がこれまで専門として行ってきたトランザクショノグラフィが適用できるのである。それに加えて，もちろん『延喜式』には財務記録以外にも官僚組織の構造などの記述が非常に多いため，一次的な文字資料として付帯情報を加えながらテクストを構造化することによって，検索利便性の高いデータベースの基盤を形成することも有意義である。関連する先行プロジェクトとして，カリフォルニア大学バークレー校の提供する Japanese Historical Text Initiative では，延喜式の 1 ～ 10 巻までの本文と英訳，対応する画像をウェブ上で閲覧できるようになっている[(12)]。一方で本プロジェクトは，延喜式全 50 巻を対象にマークアップを行い，利用者の研究関心に応じたデータ提供を目指すものである。

1.3 データベース構築の作業フロー

以上のように，延喜式の内容に踏み込んだ充実したデータベースを構築するには，例えば次のよ

うな作業フローが必要となる。

①テクストの選定

② TEI に準拠したベーステクストの作成

③ベーステクストを基に，細かいテクストの解釈をデータに反映

④人力でのデータチェックやミスの修正

本節では，上記のフローに沿って[13]，作業内容の概要を述べていきたい。①まず本プロジェクトでマークアップの対象としているテクストは，歴博所蔵の土御門本であり，校訂テクストの元データとしては集英社版のものを用いている。

巻	式名	条		【巻】	【頁】	標目	本文
1	四時祭上	1	四時祭式上	上	22		四時祭上
1	四時祭上	1	大中小祀	上	22		凡践祚大嘗祭為大祀，祈年・月次・神嘗・新嘗・賀茂等祭為中祀，大忌・風神・鎮花・三枝・相嘗・鎮魂・鎮火・道饗・
1	四時祭上			上	22		薗・韓神・松尾・平野・春日・大原野等祭為小祀，〈風神祭已上，並諸司斎之，鎮花祭已下，祭官斎之，但小祀祭官斎者，内裏不斎，其遣勅使之祭者斎之，〉
1	四時祭上	2	祭日	上	22		凡祈年祭二月四日，大忌･風神祭並四月･七月四日，月次祭六月･十二月十一日，神嘗祭九月十一日，
1	四時祭上			上	22		其子・午・卯・酉等日祭，各載本条，自余祭不定日者，臨時択日祭之，
1	四時祭上			上	22		二月祭
1	四時祭上	3	祈年祭	上	22		祈年祭神三千一百卅二座
1	四時祭上			上	22		大四百九十二座〈三百四座案上官幣，一百八十八座国司所祭，〉
1	四時祭上			上	22		小二千六百卌座〈四百卅三座案下官幣，二千二百七座国司所祭，〉

図 1　集英社版の校訂テクストに見る延喜式の資料群としての構造

マークアップ 1　TEI で作成した延喜式のベーステクスト例

```
<div type=" 式 " subtype=" 条 " n="1" corresp=" 四時祭上 ">
    <head><title corresp="1 四時祭式上 " n=" 上 _22"/> 四時祭上 </head>
    <div type=" 条 " n="1.1" corresp=" 四時祭上 ">
        <p><title corresp=" 大中小祀 " n=" 上 _22"/> 凡践祚大嘗祭為大祀，祈年・月次・神嘗・
        新嘗・賀茂等祭為中祀，大忌・風神・鎮花・三枝・相嘗・鎮魂・鎮火・道饗・薗・韓神・松尾・
        平野・春日・大原野等祭為小祀，〈風神祭已上，並諸司斎之，鎮花祭已下，祭官斎之，但小
        祀祭官斎者，内裏不斎，其遣勅使之祭者斎之，〉 </p></div>
    <div type=" 条 " n="1.2" corresp=" 四時祭上 ">
        <p><title corresp=" 祭日 " n=" 上 _22"/> 凡祈年祭二月四日，大忌・風神祭並四月・七
        月四日，月次祭六月･十二月十一日，神嘗祭九月十一日，其子･午･卯･酉等日祭，各載本条，
        自余祭不定日者，臨時択日祭之，二月祭 </p></div>
   <div type=" 条 " n="1.3" corresp=" 四時祭上 ">
       <p><title corresp=" 祈年祭 " n=" 上 _22"/> 祈年祭神三千一百卅二座　大四百九十二
        座〈三百四座案上官幣，一百八十八座国司所祭，〉小二千六百卌座〈四百卅三座案下官幣，
        二千二百七座国司所祭，〉 </p></div>
<!-- 途中省略 -->
</div>
```

②次に，TEI に準拠したベーステクストの作成にあたっては，延喜式が持つ「資料群としての構造」を表現することを目的とした。すなわち，延喜式は全 50 巻の資料群であるが，図 1 に示したように集英社版の校訂テクストを見てみると，巻 / 式 / 条という階層構造を持ったテクスト群であることがわかる。それぞれの条文が上・中・下巻のどのページに位置するかという情報もある。

これらの資料群としての構造をもとに，TEI に準拠したベーステクストを作成すると，**マークアップ 1** のようになる。ベーステクストというのは，テクストの内容を細かくデータ化していくというよりは，テクストとしてどのような構造を持っているか（行や段落など），という情報をデータ化したものであると理解されたい。TEI では，構造上のあるまとまりを タグで，ひとつの段落を <p> タグで表現することが多いため，延喜式のテクスト構造からして，一つの条文をひとまとまりとして， や <p> タグでマークアップすることとした。

このように TEI では，基本的な用途を想定して用意されている 500 以上のタグの中から[(14)]，自身のプロジェクトに沿うデータセットを選定することが一般的である。実際のデータ化の過程では，8000 行におよぶ延喜式の全テクストを対象に手入力でマークアップをするのは現実的でないので，ベーステクストの作成についてはもともと存在した行や条番号の情報をもとに自動的に処理した。

③ベーステクストを作成した後は，テクストの内容に踏み込んで細かいデータ化を行う。「モノの移動」をデータ化するトランザクショノグラフィには，まずモノをマークアップしておく必要があるので，その例も簡単に示しておきたい。

マークアップ 2 は，場所の名前とモノの情報，そして原文で割書きになっている箇所のマークアップを行った例である。このうち，場所の名前や割書きの箇所については，事前知識として共有されている部分があるとともに，割書きに関してはあらかじめ校訂テクストの中に〈 〉という記号で囲まれている部分が該当していたため自動的な処理を行った。しかし，モノのタグ付けとなるといささか厄介である。これについて次のフローで説明する。

④ここまでの TEI データ作成過程を見てみると，かなりの部分はプログラミングによる自動化処理が適用できることがわかる。では，モノの記述に関してはどうだろうか。実は延喜式には，「酒一斗」などのように，「品目」→「数量（漢数字）」→「単位」という順でモノが記述されていることが多い。このような法則が見つけられると，プログラミングによる自動化処理も適用させやすい。しかしながら，「春一日」などのように，漢数字で数量が記述されながらも，モノを表す記述でないこともしばしばである。

マークアップ 2　テクストの内容に踏み込んだデータ化の例

```
<div type=" 条 " n="24.8" corresp=" 主計上 ">
  <p><title corresp="山城国" n="中_854"/><placeName xml:id="山城国">山城国
    </placeName> 調, <measure xml:id="調 _山城" commodity="広席" quantity="280"
        unit="枚">広席二百八十枚</measure>，狭席五百九十枚，折薦八百五十八枚，葉薦
        四百六枚, 食薦一千五百枚, <note type="割書"> 随時損益, 余国准此 </note> 自余輸銭,
        </p></div>
```

そこで本プロジェクトでは，「品目」「数量」「単位」のパターンで現れる記述をすべて抽出し，それぞれについてモノの記述となっているかどうかの判定を人力で行った。この作業は膨大で，歴博の清武雄二氏のご助力を得て，4000項目のデータを目視で確認していただいた。

本プロジェクトは，①〜④で見た作業フローのように，ある程度まではプログラミングによる自動化処理に基づいてデータを作成し，データのチェックやミスの修正にあたって専門的に検証している。これは，人文学のための研究基盤としてのデータ構築のプロセスにおける，人文学研究者とエンジニアの共同作業のあり方として現実的だろう。

本章の最後に，プロジェクトの主な成果について言及しておきたい[15]。

本プロジェクトの研究成果の一部は，2017年11月にカナダのヴィクトリア大学で開催されたTEI年次国際大会のポスター発表として公開された[16]。

内容としては，延喜式に出現する度量衡の記述をマークアップするためのデータセットを提案した。すなわち，古代日本における度量衡は，斤・両・分・銖など重量の単位に典型的に見られるように，十進法以外に基づく換算の体系も有していたため，それらの数量や単位を原資料の記述のまま構造化できるようなTEIエレメントおよび属性の必要性を指摘したのである。

実際のポスター発表およびその後のTEIコンソーシアムのオンライン上の議論を経て[17]，新たなデータセットがTEIガイドラインに採択されることとなり，さまざまな文化圏の歴史資料において多種多様な度量衡のあり方を構造化できる展望が開けた。この事例は，欧米圏を中心に開発・整備が進められてきた枠組みを東アジア文化圏で批判的に導入し，その上で文化的特徴に基づくフィードバックを行うことにより，TEIの収める射程がより国際的に広がったという意義を有する。

このような作業を行い，延喜式のデータ構築を実施した。次に，これらのデータ構築がもたらす意義をTEIマークアップによる成果のみならず，プロジェクト全体がもたらす意義について述べることとしたい。

❷……人文情報学のありようとしてのTEI基礎データ構築の意義と課題

ここまで述べてきたように，本プロジェクトにおいて，延喜式のTEIデータができる意義は非常に大きいものである。TEIのテクストそのものを作ることの意義だけではなく，日本資料のTEI構築事例がTEIのプロジェクト全体にも貢献しうることを述べてきた。そして，さらに延喜式を例とすることで，下記のメリットがあげられる。延喜式は神祇式のような文章体で書かれたものと，主計式を代表とするような，「帳簿」のように書かれたものの二種類がある。この帳簿については，小風が行ったようなマークアップが可能であり，文章体である場合には，漢文のマークアップ事例となる。この点において，日本の漢文資料の様々なモデルとなる可能性を秘めている。したがって，今後の日本のTEIプロジェクトのスタートアップとしては，適正な資料ではないかと考える。

また，本プロジェクトは，延喜式の現代語訳・英語訳と並行して進められているため，データの流通や，多言語への対応の事例としても行いやすい点がメリットであるといえよう。言語の切り替えや，条ごとの表示手法など，様々な事例を実験することができるのは，特徴であると言える。

2.1 課題としてのマニュアルとより容易なマークアップ

上記のような状況の中で，延喜式研究における TEI の有用性は一定程度見込まれるであろうことは見通しがある。しかし，これらのデータをより広範に研究として用いるためには，これらをより容易な手法でデータを作ることが求められている。これまで，日本における TEI マークアップ手法の検討の他の事例としては，永井正勝らなどの仕事がある[18]。これらの研究は，どちらかといえば，ある個別の研究目的に即したものである。永崎氏は大正新脩大蔵経データベースに TEI を適用する検討を行なっている[19]。これは基盤的なテクストデータへの TEI 適用という数少ない事例ではあるが，その基盤構築の工夫などについては，まだ共有されていない。より端的にいえば，TEI は極めて複雑なマークアップルールを持っており，それらを活用して基礎データを作るためには，もう一つハードルを持つという課題があるのである。

そこで，本プロジェクトでは，さらに，基盤となりうるデータ構築の手法そのものの共有化をはかることとした。具体的には，マークアップ作業の記録をもとに，それをマニュアルとして整備し，共有することを目指したのである[20]。

TEI それ自体が，テクストをどのように認識するかを可視化し，共有化するための手法であることは第 1 章において述べた。しかし，実際には TEI は極めて複雑であり，簡単にマークアップすることが難しいという点も事実である。

しかし，国際標準に則った基盤テクストを構築し，国際的に流通させる意義は，述べた通りであり，さらにひいては世界における日本の研究および東アジアの研究にとっても重要である。単に全文が Web 上にあるのではなく，構造化されたテクストとすることで，『延喜式』そのものを直接読まなくとも，なんらかの見当をつけるなどの活用方法も考えうる。日本を対象とする歴史研究者のみならず，中国を対象とする研究者などに対しても有益になりうる。延喜式は東アジアにおける研究価値は高い。これらの点からも，より汎用的なマークアップを施すことが必要であると考えた。

そこで，テクストをどのように理解したかを書く TEI に対し，さらにメタなレベルでのマニュアルを作成することで，TEI データを多くの人が基盤データとして作成可能にすることを目指した。さらに付け加えるなら，万一，この延喜式 TEI データがなんらかの理由で歴博から離れて管理されることになっても，このマニュアル自体が，当時どのような意図で構築されたかの記録となり，長期保存を目指した TEI にさらに長期的なメタ情報が加わり，より長期的なデータ活用につながりうると考えられる。

本 TEI マニュアルの具体的な構成は下記の通りである。

1．メタデータ記述

ここには，基本的には TEI Header を中心に説明を述べている。歴史資料について説明する必要な要素・データ作成・資料作成などに関連した人物，画像との対応付けなど，歴史的な資料をエンコーディングする際に，汎用的に必要であろうと考えられる部分に関する基礎的な説明を述べている。

2．全体構造記述：巻や章など，区切りごとに構造化することの必要性

『延喜式』の特性に応じて，どの部分にどのようなタグを付したのかを説明している。この部分は一般的には歴史資料の特性に応じて変更しなければならない部分ではある。ただし，『延喜式』には大きく2つ，もしくは3つの文の様式を持っている。一つは，祝詞などのような漢文の文章体の様式，もう一つは法令ごとに帳簿のように説明を記した帳簿様式である。さらに分けるなら，文書の例示のような「見本」様式が帳簿様式から分離できる。このように，延喜式は，複数の様式を持っており，比較的多くの歴史資料でも参照しやすい特徴がある。そのため，延喜式を例とすることで比較的汎用性の高いマニュアルとなると考えられる。

3．目的に即した記述

これは著者のうち小風が検討したものなど，関連するマークアップを記録として残したものが，現時点では入っている。ここには，『延喜式』マークアップのための個別の研究で行われたデータを蓄積する。これ自体は，必ずしも汎用的ではないが，基盤研究から発展した検討を行う際，どのようなことができるのかの参照を行うことを目指している。今後も，小風以外にも『延喜式』のマークアップを用いた個別研究事例をここに蓄積する予定である。

また，これ自体は研究の記録としても機能し，一つの歴史資料に対してどのような研究が行われたのか，実際に流通しているマークアップデータはどのような意図で作られたのかを残すものでもある。このようなデータは，一義的には論文で記述されるが，それをより具体化したものをここに残す。

4．多言語対応

『延喜式』のプロジェクトにおいては，条文を英訳することも検討されている。英訳の作業自体は，現時点では中途であるため，まだマニュアル上では記述されていない。今後の課題となっている。

5．スキーマ

TEIのデータをどのようにカスタマイズしたかの記録である。長期的なデータ活用のためでもある。この記録があることで，国際標準に加え，どのような独自データを作っているのかが残る。国際標準のみでは限界のある分析を独自ルールで行うことがTEIは可能であるが，その独自ルールがどのようなものになっているのかが，判然としないようでは，国際標準にのっとった意義が半減してしまう。そのような問題を回避するための記録が，この部分に残ることとなる。

6．表示やアプリケーションの例

ここで作成したデータが，どのように応用されるかの例を記載している。最終的な表示方法や，アウトプットも含め，ここに記載している。

全体の構成としては上記の通りである。主に1・2で最も基本的なマークアップを可能にし，3・4・

5においてより応用度の高いものを示すという構成となっている。6はそれらの流れと少し異なり，TEIの意義を示すための機能も果たしている。

このようなマニュアルを作ることで，TEIの複雑なルールからの回避手段を少しでも増やす手法を検討した。資料の情報をよりメタなレベルで残すことで，資料データそのものをより長期に残すことが可能になる。それは，よくわからないデータはより消失の危機に晒されやすくなるが，データの意義と価値がわかることで，データをマイグレーションし，残し続けるモチベーションを高めることができるためである。なお，このことは本質的にはデジタルデータに限ったものではない。資料はそのコンテクストと意義付けを行った結果，はじめてその価値を発揮するものであるという点からは，デジタルデータも，物体としての資料も同様の「意義の継承」を行う必要がある。

このことにより，TEIによるデータの構築を日本において，より容易にする可能性が開ければと考える。一方，マニュアルの構築という対応は，本質的な解決方法ではないため，技術的な解決方法も含む検討は今後の課題となるであろう。

おわりに

以上，延喜式におけるTEIの構築の意義と，基盤となるテクストデータ構築の重要性について述べてきた。本論文の最後に，日本の歴史資料の流通の重要性について改めて述べて終わることとしたい。

現在，日本においては，「ジャパンサーチ」を代表として，日本の歴史・文化に関するデジタル情報の発信への動きは広く行われている(21)。そのこと自体は大変に望ましく，これまでに遅れをとってきたとされる日本の歴史文化資料が，オープンな形で広く流通すれば，日本に関する研究が国際的に進む重要な基盤となるであろう。しかし，その中にはこのようなテクストデータは決して多く存在しない。国際的な流通という観点からすると，テクストデータを日本語で作ったとしても，言語障壁に阻まれてしまうのではないか，という懸念もあるのであろう。しかし，本事例でも述べたように，例え日本古代のテクストであっても，東アジア全体での検討材料へと発展しうる可能性もある。また，日本以外で日本研究を行うためには，このようなWebの情報は極めて重要な位置を占めることになるであろう。

日本史研究の活性化という観点からは，このような歴史資料や研究手法の可視化は欠かすことができない。人文学や歴史学が「危機」と呼ばれる現在であるからこそ，基盤データを構築し，自由に流通し，様々な可能性を開く研究を検討することが求められるのではないだろうか。

註及び参考文献

（1）――東京大学史料編纂所　SHIPS　wwwap.hi.u-tokyo.ac.jp/ships/db.html（閲覧日2018年9月1日。以下，註2を除き同じ）
（2）――東寺百合文書WEB http://hyakugo.kyoto.jp/（2019年1月15日確認）
（3）――国立公文書館デジタルアーカイブ　https://www.digital.archives.go.jp/
（4）――国文学研究資料館「新古典籍総合目録」https://kotenseki.nijl.ac.jp/
（5）――「みんなで翻刻」https://honkoku.org/

（6）――大正新脩大蔵経データベース（SAT）http://21dzk.l.u-tokyo.ac.jp/SAT/
（7）――TEI（Text Encoding Initiative）http://www.tei-c.org/
（8）――トランザクショノグラフィについては，次を参照されたい。Kathryn Tomasek and Syd Bauman, 'Encoding Financial Records for Historical Research', *Journal of the Text Encoding Initiative*［Online］, Issue 6；December 2013, URL：http://jtei.revues.org/895［拙訳「歴史研究のため財務記録史料マークアップ手法」東京大学術機関リポジトリ，2015 年 6 月。http://hdl.handle.net/2261/56940］
（9）――ワークショップの発表詳細については，次を参照のこと。http://medea.hypotheses.org/
(10)――小風尚樹・永崎研宣・下田正弘・A. Charles Muller「歴史的商取引叙述のための TEI 拡張モデルに基づくマネーフロー可視化と多言語史料分析のためのインタフェース構築：レイ・オズボーン艦隊事件を手がかりに」『情報処理学会研究報告．人文科学とコンピュータ研究会報告』2016–CH–110(8)，2016年5月，1–6頁。http://id.nii.ac.jp/1001/00159412/
(11)――虎尾俊哉『延喜式』吉川弘文館，初出 1964 年
(12)――https://jhti.berkeley.edu/Engi%20shiki.html
(13)――TEI に準拠したデータの作成には，どの程度人的・時間的コストを割くかという問題があり，その割けるコストによって，データの充実度が変わってくる。作業フローの②に示したように，ベーステクストを作成することで留めたとしても，プロジェクトの目的と状況によっては十分な成果となる。プロジェクトの状況に応じて，どの程度充実した TEI データを作成すれば良いか，というガイドラインについては以下が参考になる。Kevin Hawkins and Michelle Dalmau, eds., 'Best Practices for TEI in Libraries: A guide for mass digitization, automated workflows, and promotion of interoperability with XML using the TEI', 2017 November. http://www.tei–c.org/SIG/Libraries/teiinlibraries/
(14)――TEI では，テクストの性質に応じて，様々なデータセットを用意している。詳しくは，TEI ガイドラインを参照されたい。http://www.tei–c.org/release/doc/tei-p5–doc/ja/html/index.html
(15)――本節の記述については，以下に基づいている。小風「The 2017 Annual Meeting of the TEI Consortium に参加して」『アビナヴァトリピタカ［科研基盤 S ニューズレター「仏教学新知識基盤の構築」代表者：下田正弘］』2018 年 3 月。
(16)――Naoki Kokaze, Kiyonori Nagasaki, Makoto Goto, Yuta Hashimoto, Masahiro Shimoda, and A. Charles Muller, 'TEI/XML Methodological Examination on Unit Conversion not Based on the Metric System', The 2017 Annual Meeting of the TEI Consortium, Victoria, British Columbia, Canada, November 2017. https://hcmc.uvic.ca/tei2017/abstracts/t_107_kokazeetal_unitconversion.html
(17)―― 'How to encode measurement', opened by naoki_kokaze, https://github.com/TEIC/TEI/issues/1707
(18)――高橋洋成，永井正勝，和氣愛仁，画像，TEI，LOD を用いた文字研究・言語研究のためのプラットフォームの構築，情報処理学会研究報告　2015–CH–105(5)，pp1–8.
(19)――永崎研宣　仏教文献のための構造的なデジタルテクストの記述と活用．印度學佛教學研究 63(2)：1094－088，2015.
(20)――図書館向けのマークアップマニュアルとしては以下のようなものもあるが，さらに研究機関向けのものを目指す。http://www.tei-c.org/SIG/Libraries/teiinlibraries/3.1.0a/main-driver.html
(20)――ジャパンサーチ https://jpsearch.go.jp

小風尚樹（東京大学大学院人文社会系研究科大学院生，国立歴史民俗博物館共同研究研究協力者）
後藤 真　（国立歴史民俗博物館研究部）

（2018 年 9 月 18 日受付，2019 年 3 月 28 日審査終了）

Bulletin of the National Museum of Japanese History
Vol. 218 December 2019

Application of TEI to *Engishiki* and Japanese History Text Data Sharing

KOKAZE Naoki and GOTO Makoto

In this paper, we have examined the digitalization of the text information of *Engishiki* and strategies for its distribution. In particular, we have discussed the data obtained through the application of an international standard known as TEI (Text Encoding Initiative), and introduced ways to more broadly circulate the text data of Japanese historical materials.

This paper we do not discuss in detail the technical aspects of our project, but we rather focus on its significance from the standpoint of historical research. Aiming to facilitate the practical use of the text data for international circulation and research, we have applied TEI markup to the *Engishiki*. TEI (Text Encoding Initiative) is a project intended to promote the international circulation and use of textual data in the humanities, and it's also called the standards developed within these project. Since TEI is a tool specifically developed for the digital humanities, it offers the advantage of allowing users to record information on one's understanding of historical materials through the use of "elements" (pieces of information to which a tag has been attached). Bearing such merits in mind, the authors have proposed a digitalization model based on TEI.

In particular, we have applied TEI markup to quantitative data of *Engishiki* and ventured into an examination of the possibility of parsing an overall picture of the text through the use of a method known as "transactionography." In addition, we have also discussed ways of creating a manual on how to use these markups as information bases.

The development of methods of visualization for historical materials and research methodologies is necessary in order to more active the research on Japanese history. At a time in which the humanities and history are said to be in "corner", it is all the more urgent to construct base information, distribute it freely, and devise new methods that can lead to new avenues for inquiry.

Keywords: TEI (Text Encoding Initiative), markup to the *Engishiki*, Information infrastructure construction, Historical informatics, the digital humanities