



Journal of the Text Encoding Initiative

Issue 12 | July 2019 - May 2020

Selected Papers from the 2017 TEI Conference

Toward a Model for Marking up Non-SI Units and Measurements

Naoki Kokaze, Kiyonori Nagasaki, Makoto Gotō, Yuta Hashimoto, A. Charles Muller and Masahiro Shimoda



Electronic version

URL: <http://journals.openedition.org/jtei/1996>

DOI: 10.4000/jtei.1996

ISSN: 2162-5603

Publisher

TEI Consortium

Electronic reference

Naoki Kokaze, Kiyonori Nagasaki, Makoto Gotō, Yuta Hashimoto, A. Charles Muller and Masahiro Shimoda, « Toward a Model for Marking up Non-SI Units and Measurements », *Journal of the Text Encoding Initiative* [Online], Issue 12 | July 2019 - May 2020, Online since 29 June 2019, connection on 26 June 2020. URL : <http://journals.openedition.org/jtei/1996> ; DOI : <https://doi.org/10.4000/jtei.1996>

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.

Toward a Model for Marking up Non-SI Units and Measurements

Naoki Kokaze, Kiyonori Nagasaki, Makoto Gotō, Yuta Hashimoto, A. Charles Muller,
and Masahiro Shimoda

ABSTRACT

This paper presents a markup model for encoding non-SI units and measurements. Historical texts contain many examples of compound measurements, composed of sets of units and numerical components. Instead of using the `<measure>` element, which requires a single set of `@unit` and `@quantity`, we propose a newly defined set of tags for encoding idiosyncratic measurement semantics, namely `<unitDecl>` (`model.encodingDescPart`), `<unitDef>` (`model.global` and contained by `<unitDecl>`), `<unit>` (`model.measureLike`), and a relevant attribute `@factor` (which shows factors of numerical values given in a referenced `<unit>` element). All of these elements and attributes will be included in the TEI P5 Guidelines, and they are especially useful when encoding

units that are not based on the decimal system. Though this paper offers example encodings based on a Japanese historical source, the *Engi-Shiki*, this model is also applicable to the markup of units used for measurement within various cultural spheres other than Japan.

INDEX

Keywords: non-SI units and measurements, TEI ODD, TEI customization, Engi-Shiki, factor attribute

ACKNOWLEDGEMENTS

We would like to thank all the researchers who engaged in the discussion to develop the markup model, both at the 2017 Annual Meeting of the TEI Consortium and on the TEI Consortium GitHub site, especially Elisa Beshero-Bondar, Dario Kampkaspar, Duncan Paterson, Laurent Romary, and James Cummings. Kathryn Tomasek deserves the greatest acknowledgment for her encouragement and for inviting us to join the research project MEDEA (Modeling semantically Enriched Digital Edition of Accounts).

The project reported in this paper is funded by the National Institute of Japanese Literature and the National Museum of Japanese History, and supported by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (S), “Construction of a New Knowledge Base for Buddhist Studies: Presentation of an Advanced Model for the Next Generation of Humanities Research,” represented by Professor Masahiro Shimoda at the University of Tokyo.

1. Introduction

1.1 Research Aim

- 1 This paper presents a markup model for encoding units and measurements which are not based on the International System of Units (SI). The TEI P5 Guidelines recommend the use of the `<measure>` element, which requires a single `@unit` and `@quantity` (TEI Consortium 2018, “`<measure>`”)¹ for marking up the units used for measurement. But the main issue we address in this paper is that some historical units and measurements cannot be encoded in a single `<measure>` element because they are compounds containing a set of variable units and numerical components. A

typical example of the compound unit would be a British imperial length measurement, such as 3 yds 2 ft 11 in (3 yards, 2 feet, and 11 inches). The same issue arises with the Japanese historical measurements, as we can see from examples in the *Engi-Shiki*, which was compiled as an administrative manual in ancient Japan around the tenth century CE (Bock 1970). Figure 1 provides an example of the compound units used to measure the weight of copper. In this text, there are variable units, namely 斤 kin, 兩 ryō, 分 bu, and 銖 shu. They were not based on the decimal system, since 1 斤 kin was equal to 16 兩 ryō, 1 兩 ryō was 4 分 bu, and 1 分 was 6 銖 shu. In this example, “銅二千五百十六斤十兩二分四銖 (Copper 2516 kin 10 ryō 2 bu 4 shu)” shows a specific weight expressed using these four different units.

Figure 1. Sample description of a compound measurement for copper in the *Engi-Shiki*.



National Museum of Japanese History.

- 2 In order to encode such compound units, we propose three new elements and one new attribute: `<unitDecl>` (`model.encodingDescPart`), `<unitDef>` (`model.global` and contained by `<unitDecl>`), `<unity>` (`model.measureLike`), and a relevant attribute `@factor` (which shows factors of numerical values given in a referenced `<unit>` element), so that we can clarify the relations among the units in the `<teiHeader>`. All of these will be included in the TEI P5 Guidelines in due course, and the ODD File for them will be described in detail later in this paper.

1.2 Further Implications

- 3 Although this paper provides a markup model based on a Japanese historical source, this model also enables us to encode idiosyncratic systems of measurement within various cultural spheres other than Japan. This set of tags is useful for encoding units that are not based on the decimal system or whose unit conversion system changed over time. For instance, conversion relations among the English coal measures keel, chalders, and bushel had changed between the fifteenth century and the implementation of the Imperial Weights and Measures Act of 1824 (figure 2).

Figure 2. Pre-Imperial English coal measures (Zupko 1977, 144, 151).

<i>British Pre-Imperial Units</i>		
English Coal Measures		
bushels	chalders	keels
1421–1676		
1 32 640	1 20	1
1676–1824		
1 36 576	1 16	1

The sample encoding of these units will be provided in [section 4.2.](#) below.

- 4 Since texts related to measurement indicate such regional and chronological differences, it is necessary to mark up the original historical figures as precisely as possible. The new markup model enables us to express the complex and changing semantics of the units and measurements.

2. Project-based Discussion

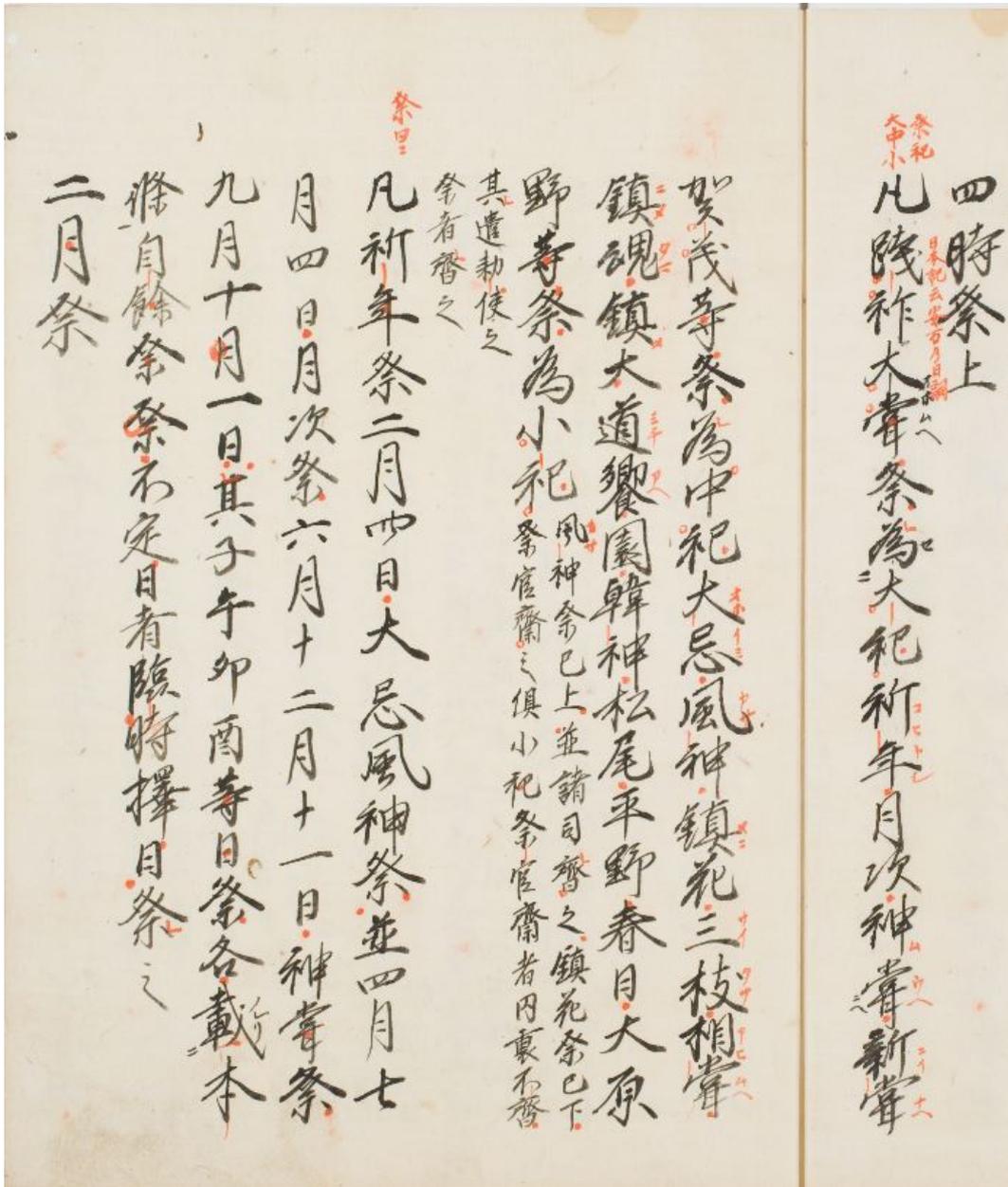
2.1 Our Project

- 5 In this section, we describe the outline of our project—marking up the texts of the *Engi-Shiki*—and compare it with other related projects and previous studies.
- 6 Delmer Brown, an American scholar of the *Engi-Shiki*, describes it as “a 50-volume work compiled between ... 907 and 927 [CE]. The first 10 volumes are Imperial Shinto regulations (jingi 神祇) and the last 40 are codifications of criminal (ritsu 律) and administrative (ryō 令) law” (UC Berkeley 2010). According to one of the most influential researchers on ancient Japanese history, Toshiya Torao, the *Engi-Shiki* contains a wide range of detailed regulations on society and administration during the period from the Nara to the Heian era (eighth to thirteenth centuries). For instance, the text lists rituals and festivals held in various parts of Japan, the designation of offerings made at rituals, tributes and taxes paid to the Ritsuryō government, and allocation of Shōzei and Kugai (rice plants used as a kind of fund which were distributed to each administrative county). To put it simply, by consulting the *Engi-Shiki*, we can grasp various aspects of both state administration and daily life in ancient Japan (Torao 1995).
- 7 But there is a historiographical problem. Because so many elements of interest are described in such rich detail, previous research based on the *Engi-Shiki* has tended to be subdivided into very specialized and partial studies. For example, Sakamoto (1979–1980) comprehensively lists the products of each administrative county; Fukushima (1971) and Ōsumi (1996) focus on the aquatic products; Satō (2012) examines the dairy products; papers and paper manufacturing are the focal point of Ōkawa and Masuda (1981); and Miyahara (2014) traces the records related to abalone, which was the most circulated food at that time. This subdivision and specialization makes it difficult to verify quantitative analyses, because there is no text database in which researchers could search for numerical values attached to specific products and items.

- 8 Our research project has been working to provide a possible solution to this problem. One of our aims is to formulate a text database enabling historians, according to their individual interests, to extract numerical values by making use of various machine-processable data on society and administration in ancient Japan. One similar research project, the [Japanese Historical Text Initiative \(JHTI\)](#), based at the Center for Japanese Studies at the University of California, Berkeley, offers digitized text of the first ten volumes of the *Engi-Shiki* text in Japanese and in English translation along with page images of the manuscript volumes ([UC Berkeley 2010](#)). Our project will provide TEI-based text and numerical values attached to the various kinds of products and items, comprehensively covering all fifty volumes of the *Engi-Shiki*. For this purpose, we have proposed expanding the functionality of the <measure> element ([Kokaze and Nagasaki 2016](#)).²

2.2 Basic Structure of the *Engi-Shiki*

- 9 We have based our project on the 土御門 Tsuchimikado version of the *Engi-Shiki* ([Kokuritsu Rekishi Minzoku Hakubutsukan Kanzō Shiryō Henshūkai 2000-](#)).

Figure 3. Sample passages from the *Engi-Shiki*, from volume 1 四時祭式 Shijisai-Shiki.

National Museum of Japanese History.

- 10 The basic structure of the *Engi-Shiki* has a simple hierarchy which goes down from 式 shiki to 条 jou. So, we mark up the sample passage of the *Engi-Shiki*, shown in figure 3, using nested <div> elements, as follows:

Example 1. Basic structure of the *Engi-Shiki* encoding.

```

<body>
  <div ana="四時祭上" n="1" type="式" subtype="条">
    <head type="式名">四時祭上</head>
    <div type="条" n="1.1">
      <head type="条名">祭祀大中小</head>
      <p>凡踐祚大嘗祭為大祀、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・菌・韓神・松尾・平野・春日・大原野等祭為小祀、〈風神祭已上、並諸司齋之、鎮花祭已下、祭官齋之、但小祀祭官齋者、内裏不齋、其遣勅使之祭者齋之、〉 </p>
    </div>
    <div type="条" n="1.2">
      <head type="条名">祭日</head>
      <p>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、臨時摺日祭之、二月祭</p>
    </div>
    <!-- the rest of the contents -->
  </div>
</body>

```

Table 1. Descriptions of the basic structural markup.

Element	@type	Description
div	式	Nests a <head> element and several <div> elements; @ana and @n provide the name and the index number of the 式; @subtype="条" makes the hierarchical relation between 式 & 条 clear
head	式名	Includes the 式 Shiki's Name
div	条	Nests a <head> element and a <p> element; @n gives the index number of the 条 within the 式, using "." as the delimiter (e.g., n="1.2")
head	条名	Includes the 条 Jou's Name

p		Contains the main text of each 条 Jou
---	--	--------------------------------------

- 11 Marking up the basic structure of the *Engi-Shiki* in this way allows an overview of the tree structure of the source, using the Oxygen XML Editor’s “Grid” mode, for example.

3. Markup Examples Based on TEI P5

3.1 Defining Measurement in the <teiHeader>

- 12 Having outlined the basic document structure of the *Engi-Shiki*, we can return to the encoding of the definitions of non-SI units. Needless to say, TEI provides tags for taxonomies and classification (TEI Consortium 2018, “<classDecl>”).³ If you want to categorize and define non-SI units, you can express them by the descendent elements of <encodingDesc>, as follows:

Example 2. A taxonomy of non-SI units.

```
<encodingDesc>
  <classDecl>
    <taxonomy>
      <bibl>units of measuring weight</bibl>
      <category xml:id="斤">
        <catDesc>斤is a unit measuring weight and equal to 16兩.</catDesc>
      </category>
      <category xml:id="兩">
        <catDesc>兩is a unit measuring weight and equal to 4分.</catDesc>
      </category>
      <category xml:id="分">
        <catDesc>分is a unit measuring weight and equal to 6銖.</catDesc>
      </category>
      <category xml:id="銖">
        <catDesc>銖is a unit measuring weight.</catDesc>
      </category>
    </taxonomy>
  </classDecl>
</encodingDesc>
```

- 13 In this way, relations among non-SI units of measurement might be described in a human-readable way, within the `<catDesc>` element. However, it is difficult to express the unit conversion system among the measurements in a machine-readable way, especially when those units are not based on the decimal system.

3.2 Inline Markup

- 14 Instead of defining the unit conversion system in the `<teiHeader>`, we could encode the compound measurements in the form of inline markup, as Cummings and Wilcox (2013, sec. 7) show in this example of encoding a value in old English money:⁴

Example 3. Old English money, as encoded by Cummings and Wilcox (2013).

```
<seg type="fee" rend="roman-numerals aligned-right">
  <num type="totalPence" value="1240">
    <!--orig: vli iijs iiijd -->
    <num type="poundsAsPence" value="1200">v<hi rend="superscript">li</hi></num>
    <num type="shillingsAsPence" value="36">iiij<hi rend="superscript">s</hi></num>
    <num type="pence" value="4">iiij<hi rend="superscript">d</hi></num>
  </num>
</seg>
```

- 15 One of the important points of their markup is the use of nested `<num>` elements with `@type` attributes, indicating “vli iijs iiijd” as a compound unit, not as separated units. Although this example is inspiring for our project, we would prefer not to store converted numerical values in `@value` attributes; we would rather keep the original data when it comes to inline markup. The reason for this is that the unit conversion system has altered over time, as in the case of the British coal measures shown in figure 2.

4. Markup Model

4.1 Elements and Attributes for Expressing the Metrological Semantics

- 16 In order to structure, in the <teiHeader>, a wide variety of measurement semantics within each of the historical documents, we propose a new set of elements and attributes. To take an example from the *Engi-Shiki*, the following is a definition of 斤, 兩, 分, and 銖, all of which are units for measuring weight. For the sake of readability, we omit the compulsory <fileDesc> element. Customized elements and attributes are <unitDecl>, <unitDef>, <unit>, and @factor.

Example 4. Proposed usage of <unitDecl>.

```
<encodingDesc>
  <unitDecl>
    <unitDef xml:id="斤" type="weight">
      <label xml:lang="ja">斤</label>
      <label xml:lang="ja-Latn">kin</label>
      <unit ref="#兩" factor="16"/>
      <desc xml:lang="en">斤is a unit measuring weight and equivalent to 16兩.</desc>
    </unitDef>
    <unitDef xml:id="兩" type="weight">
      <label xml:lang="ja">兩</label>
      <label xml:lang="ja-Latn">ryō</label>
      <unit ref="#分" factor="4"/>
      <desc xml:lang="en">兩is a unit measuring weight and equivalent to 4分.</desc>
    </unitDef>
    <unitDef xml:id="分" type="weight">
      <label xml:lang="ja">分</label>
      <label xml:lang="ja-Latn">bu</label>
      <unit ref="#銖" factor="6"/>
      <desc xml:lang="en">分is a unit measuring weight and equivalent to 6銖.</desc>
    </unitDef>
    <unitDef xml:id="銖" type="weight">
      <label xml:lang="ja">銖</label>
      <label xml:lang="ja-Latn">shu</label>
      <desc xml:lang="en">銖is a unit measuring weight.</desc>
    </unitDef>
```

```

</unitDecl>
</encodingDesc>

```

- 17 Here, the `<encodingDesc>` contains a `<unitDecl>` element that declares information about the measurement semantics in that source, and nests `<unitDef>` as the compulsory child. In this `<unitDef>` element, we can define the identifier and the purpose of a unit by using `@xml:id` and `@type` attributes respectively. The `<label>` and `<desc>` elements contained by the `<unitDef>` can be used to describe how a unit of measurement is used in the source and give some information about that unit. Inside the `<unitDef>` we find the `<unit>` element, which was originally suggested by Laurent Romary on the TEI Consortium GitHub repository.⁵ Though we have benefited from his contribution in developing our markup model, his approach deals mainly with examples of in-line markup within the `<body>` element. However, the proposed `<unit>` element can reasonably be used in the `<teiHeader>` to indicate the relationship of the conversion between units, along with `@factor`, which shows the numerical formula. Together, these elements enable us to mark up units that, in their original forms surviving in the historical financial records, are not based on decimal notation.

4.2 ODD File

- 18 We propose the following TEI ODD specifications⁶ for marking up units:

Example 5. Proposed ODD specification for marking up units.

```

<elementSpec ident="unit" ns="http://www.example.org/ns/nonTEI" mode="add">
  <desc>contains a symbol, a word or a phrase referring to a unit of measurement
in any kind of formal or informal system.</desc>
  <classes>
    <memberOf key="model.measureLike"/>
    <memberOf key="att.canonical"/>
    <memberOf key="att.global"/>
    <memberOf key="att.typed"/>
  </classes>
  <content>
    <macroRef key="macro.paraContent"/>
  </content>
  <attList>

```

```

    <attDef ident="factor" mode="add">
      <desc>shows factors of numerical values given in a referenced <gi>unit</gi>
element</desc>
      <datatype>
        <dataRef key="teidata.numeric"/>
      </datatype>
    </attDef>
  </attList>
</elementSpec>
<elementSpec ident="unitDef" ns="http://www.example.org/ns/nonTEI" mode="add">
  <desc>contains descriptive information related to certain unit.</desc>
  <classes>
    <memberOf key="model.global"/>
    <memberOf key="att.global"/>
    <memberOf key="att.dataable"/>
    <memberOf key="att.canonical"/>
    <memberOf key="att.typed"/>
  </classes>
  <content>
    <alternate maxOccurs="unbounded">
      <elementRef key="unit" minOccurs="1"/>
      <classRef key="model.labellike"/>
    </alternate>
  </content>
</elementSpec>
<elementSpec ident="unitDecl" ns="http://www.example.org/ns/nonTEI" mode="add">
  <desc> (unit declarations) provides information about non-SI (the International
System of Units) units and measurement.</desc>
  <classes>
    <memberOf key="model.encodingDescPart"/>
    <memberOf key="att.global"/></classes>
  <content>
    <elementRef key="unitDef" minOccurs="1" maxOccurs="unbounded"/>
  </content>
</elementSpec>

```

- 19 Among these definitions in this ODD File, we need to explain why we choose `att.datable` in the `<unitDef>` element. In order to illustrate the reason for this choice, we provide a sample encoding of the conversional transition within the English coal measures mentioned in [section 1.2](#) above:

Example 6. Encoding of the conversions relating to English coal measures.

```

<encodingDesc>
  <unitDecl>
    <unitDef xml:id="keel_prev" type="coal_measure" from="1421" to="1676">
      <label xml:lang="en">keel</label>
      <unit ref="#chalder_prev" factor="20"/>
      <desc>In England, keel was a unit for measuring coal and equal to 20 chalders
from 1421– 1676.</desc>
    </unitDef>
    <unitDef xml:id="chalder_prev" type="currency">
      <label xml:lang="en">chalder</label>
      <unit ref="#bushel_prev" factor="32" from="1421" to="1676"/>
      <unit ref="#bushel_new" factor="36" from="1676" to="1824"/>
      <desc>In England, chalder was a unit for measuring coal and equal to 32
bushels from 1421–1676 and equivalent to 36 bushels from 1676–1824.</desc>
    </unitDef>
    <unitDef xml:id="bushel_prev" type="currency">
      <label xml:lang="en">bushel</label>
      <desc>In England, bushel was a unit for measuring coal.</desc>
    </unitDef>
    <unitDef xml:id="keel_new" type="coal_measure" from="1676" to="1824">
      <label xml:lang="en">keel</label>
      <unit ref="#chalder_new" factor="12"/>
      <desc>In England, keel was a unit for measuring coal and equivalent to 12
chalders from 1676–1824.</desc>
    </unitDef>
    <unitDef xml:id="chalder_new" type="currency">
      <label xml:lang="en">chalder</label>
      <unit ref="#bushel_new" factor="36" from="1676" to="1824"/>
      <desc>In England, chalder was a unit for measuring coal and equivalent to 36
bushels from 1676–1824.</desc>
    </unitDef>
    <unitDef xml:id="bushel_new" type="currency">
      <label xml:lang="en">bushel</label>

```

```

    <desc>In England, bushel was a unit for measuring coal, whose conversion
formula was changed chronologically.</desc>
  </unitDef>
</unitDecl>
</encodingDesc>

```

4.3 Markup Solutions

- 20 Finally, we present two markup examples of a passage in the *Engi-Shiki*. First we encode the passage based on the current TEI P5 Guidelines, using the `<measure>` element and the `@unit` attribute, and then we demonstrate an example using `<unit>` elements within the `<body>` texts. These two examples do not necessarily differ from each other except in encoding, but both refer to the `@xml:id` attribute defined in the `<unitDef>`.
- 21 The text shown below is a part of the section describing regulation of taxes imposed upon the two administrative counties, 備中 Bicchū and 長門 Nagato. In this case, the taxes were paid annually to the contemporary Ritsuryō government in copper 銅 and lead 鉛 mineral resources.

Example 7. Current scheme, `<measure>` with `@unit` attribute.

```

<div ana="主税上" n="26.95" type="条">
  <head type="条名">鑄錢年料</head>
  <p>
    凡鑄錢年料銅鉛者、<placeName xml:id="備中国">備中国</placeName>
    <measure commodity="銅" xml:id="備中_主税_銅" quantity="800" unit="#斤">
    銅八百斤</measure>、
    <placeName xml:id="長門国">長門国</placeName>
    <measure commodity="銅" xml:id="長門_主税_銅" type="whole">銅
    <measure quantity="2516" unit="#斤" type="each">二千五百十六斤</measure>
    <measure quantity="10" unit="#兩" type="each">十兩</measure>
    <measure quantity="2" unit="#分" type="each">二分</measure>
    <measure quantity="4" unit="#銖" type="each">四銖</measure>
    </measure>、
    <measure commodity="鉛" xml:id="長門_主税_鉛" type="whole">鉛
    <measure quantity="1516" unit="#斤" type="each">千五百十六斤</measure>
    <measure quantity="10" unit="#兩" type="each">十兩</measure>
    <measure quantity="2" unit="#分" type="each">二分</measure>
    <measure quantity="4" unit="#銖" type="each">四銖</measure>
    </measure>、
  </p>
</div>

```

```

</p>
</div>

```

22 There are some points worth mentioning in this markup example. In the `<div>` element, we provide metadata about the passage, using the relevant default TEI attributes. These attributes explain that the text contained by the `<div>` element is the 95th passage of the 26th volume of the *Engi-Shiki*, 主税上 Shuzei-no-Jō. 鑄錢年料 Chūsen Nenryō, the passage title, means the annual tax in the form of minted currency. Three `<measure>` elements that have the `@commodity` attribute contain one or more variable units and numerical components. In the second and third examples, we describe the relationship of the compound measurement, using `@type="whole"` and `@type="each"`, inspired by Cummings and Wilcox (2013). Note that in this markup example we avoid using `<measureGrp>` elements as parent elements of `<measure type="each"/>` elements, because the TEI P5 Guidelines recommend using `<measureGrp>` when marking up a group of dimensional specifications (TEI Consortium 2018, “`<measureGrp>`”).⁷

23 Or, to mark up each of the units separately, `<unit>` elements can be used in the inline markup, as follows:

Example 8. Using `<unit>` element.

```

<div ana="主税上" n="26.95" type="条">
  <head ana="鑄錢年料"/>
  <p>
    凡鑄錢年料銅鉛者、<placeName xml:id="備中国">備中国</placeName>
    <measure commodity="銅" xml:id="備中_主税_銅">銅
      <num value="800">八百</num><unit ref="#斤">斤</unit></measure>、
    <placeName xml:id="長門国">長門国</placeName>
    <measure commodity="銅" xml:id="長門_主税_銅" type="whole">銅
      <measure type="each">
        <num value="2516">二千五百十六</num><unit ref="#斤">斤</unit></measure>
      <measure type="each">
        <num value="10">十</num><unit ref="#兩">兩</unit></measure>
      <measure type="each">
        <num value="2">二</num><unit ref="#分">分</unit></measure>
      <measure type="each">
        <num value="4">四</num><unit ref="#銖">銖</unit></measure></measure>、
    <measure commodity="鉛" xml:id="長門_主税_鉛" type="whole">鉛

```

```

<measure type="each">
  <num value="1516">千五百十六</num><unit ref="#斤">斤</unit></measure>
<measure type="each">
  <num value="10">十</num><unit ref="#兩">兩</unit></measure>
<measure type="each">
  <num value="2">二</num><unit ref="#分">分</unit></measure>
<measure type="each">
  <num value="4">四</num><unit ref="#銖">銖</unit></measure></measure>
</p>
</div>

```

- 24 In this example, three `<measure>` elements that have the `@commodity` attribute contain one or more `<num>` and `<unit>` elements as sibling elements. Each `<measure>` element shows a specific weight of the mineral resource as a whole, using `@type="whole"`. We store actual values in the `@value` attribute in the `<num>` elements in the form of Arabic figures so that we can easily extract them using XSLT or other programming languages.

5. Conclusion

- 25 This paper proposes a markup model and gives examples for encoding non-SI units and measurements, based on our work with the *Engi-Shiki*. Historically, there have been many examples of compound measurement which require sets of various units and numerical components, so we have to consider how to encode them. Of course, the TEI P5 Guidelines offer us possible solutions, such as defining the taxonomies of measurement in the `<classDecl>` and providing nested `<num>` or `<measure>` elements in the form of inline markup. But the major limitation of those solutions is that they are not amenable to machine processing based on the unit conversion system.
- 26 Among the components of our proposal, `@factor` and attributes from the `att.dataable` class (e.g., `@from`, `@to`) within the `<unitDef>` element are the most effective for the purpose of describing the measurement semantics. As the example of the English coal measures keel, chalder, and bushel shows, the unit conversion formula is subject to change over time even within the same country. Therefore, it would be a reasonable practice when encoding non-SI units and measurements to describe the unit conversion formula with `@factor` between variable units, capturing the dating context with `@from` and `@to` in the `<unitDef>`.

- 27 We note in closing that our interest in encoding units of measurement is shared by other members of the TEI consortium and we look forward to further discussion.

BIBLIOGRAPHY

- Bock, Felicia G. 1970. *Engi-Shiki: Procedures of the Engi Era*. Tokyo: Sophia University.
- Cummings, James, and Pip Willcox. 2013. "Stationers' Register Online: A Case Study of a Byte-Reduced TEI Schema for Digitization (tei_corset)." *Journal of the Text Encoding Initiative* 6. <http://journals.openedition.org/jtei/926>; DOI: 10.4000/jtei.926.
- Fukushima, Yoshikazu. 1971. "Kodai Shokoku Kōnō Suisanbutsu no Bumpu ni tsuite: Sono Rekishi Chirigakuteki Kōsatsu [The Geographical Distribution of the Sea Products in the Ancient Empire of Japan]." *Jinbun Chiri [Human Geography]*, 23 (5): 495–525. DOI: 10.4200/jjhg1948.23.495.
- Kokaze, Naoki, and Kiyonori Nagasaki. 2016. "Extensibility of 'Transactionography' for Multilingual Historical Financial Records." Presented at the MEDEA Spring Meeting at Wheaton College, Norton, MA, April 7–8, 2016. Abstract available at <https://medea.hypotheses.org/478>.
- Kokuritsu Rekishi Minzoku Hakubutsukan Kanzō Shiryō Henshūkai. 2000–. *Engi-Shiki*. Kyoto: Rinsen Shoten.
- Miyahara, Takeo. 2014. *Kodai Tōgoku no Chōyō to Nōmin [Tax and peasantry in the ancient eastern counties of Japan]*. Tokyo: Iwata Shoin.
- Ōkawa, Akinori, and Katsuhiko Masuda. 1981. "Seishi ni kannsuru Kodai Gijutsu no Kenkyū [Technical study on paper making in the Nara Period (7–8th century)]." *Hozon Kagaku [Science for Conservation]* 20: 43–56. Includes an abstract in English. <http://www.tobunken.go.jp/~ccr/pdf/20/pdf/02007.pdf>.
- Ōsumi, Akiko. 1996. "Ritsuryōsei-ka ni okeru Kenkō Fukyū no Jittai: Kaisanbutsu no Kōnō Tanni wo Chūshin toshite [On the Popularization of Weights under the Ritsuryō Regime: Focusing on the Units of the Aquatic Products as Tributes]." *Shiron [Historica]* 49: 22–44.
- Sakamoto, Shintarō. 1979–1980. "Engi-Shiki kara mita Shokoku no Sambutsu Hyō [Lists of the Products of Each County in the Engi-Shiki]." *Waseda Shogaku [Waseda Commercial Review]* 281: 79–130; 284: 1–32.
- Satō, Kentarō. 2012. "Kodai Nihon no Gyūnyū/Nyūseihin no Riyō to Kōshin Taisei ni tsuite [Milk and its by-products as offerings under Japan's ancient contribution system]." *Kansai Daigaku Tōzai Gakujutsu Kenkyū Kiyō [Bulletin of the Institute of Oriental and Occidental Studies, Kansai University]* 45: 47–65.
- Tomasek, Kathryn, and Syd Bauman. 2013. "Encoding Financial Records for Historical Research." *Journal of the Text Encoding Initiative* 6. <http://journals.openedition.org/jtei/895>; DOI: 10.4000/jtei.895.

- Torao, Toshiya. 1995. *Engi-Shiki*. Vol. 8 of *Nihon Rekishi Sōsho [Series on Japanese History]*, edited by Nihon Rekishi Gakkai [The Japanese Historical Association]. Tokyo: Yoshikawa Kōbunkan.
- TEI Consortium. 2018. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 3.4.0. Last updated July 23. N. p.: TEI Consortium. <https://tei-c.org/Vault/P5/3.4.0/doc/tei-p5-doc/en/html/>.
- UC Berkeley (University of California, Berkeley). *Engi Shiki*. Japanese Historical Text Initiative. <https://jhti.berkeley.edu/Engi%20shiki.html>.
- Zupko, Ronald Edward. 1977. *British Weights & Measures: A History from Antiquity to the Seventeenth Century*. Madison: University of Wisconsin Press.

NOTES

- 1 <https://www.tei-c.org/Vault/P5/3.4.0/doc/tei-p5-doc/en/html/ref-measure.html>.
- 2 We have been greatly inspired by the projects of the [MEDEA Workshop](#), which deal with the markup modeling of historical financial records ([Tomasek and Bauman 2013](#)). We hope the future achievements of our project will be useful for their projects, as well.
- 3 <https://www.tei-c.org/Vault/P5/3.4.0/doc/tei-p5-doc/en/html/ref-classDecl.html>.
- 4 <https://journals.openedition.org/jtei/926#tocto1n7>.
- 5 “Add new element `<unit>`,” opened by @laurentromary May 20, 2016, <https://github.com/TEIC/TEI/issues/1461>.
- 6 The ODD file and encoding examples are current at the time of writing, and are under consideration by the TEI Council.
- 7 <https://www.tei-c.org/Vault/P5/3.4.0/doc/tei-p5-doc/en/html/ref-measureGrp.html>.

AUTHORS

NAOKI KOKAZE

Naoki Kokaze is a PhD candidate at the University of Tokyo, specializing in British naval history in the nineteenth and twentieth centuries. In September 2018, he joined the MA program in Digital Humanities at King’s College London, focusing on text encoding of historical sources. He has directed a cross-disciplinary workshop, Tokyo Digital History, where historians, archivists, librarians, and programmers discuss how to apply information technology to historical studies and explore digital history.

KIYONORI NAGASAKI

Kiyonori Nagasaki, PhD, is a senior fellow at the International Institute for Digital Humanities in Tokyo. His main research interest is in the development of digital frameworks for collaboration in Buddhist studies. He also investigates the significance of digital methodology in the humanities and in the promotion of DH activities in Japan. He has participated in various DH projects at several institutions in Japan and abroad. His activities also include postgraduate education in DH at the University of Tokyo, as well as administrative tasks at several scholarly societies including the Japanese Association of Indian and Buddhist Studies.

MAKOTO GOTŌ

Makoto Gotō, PhD, is an associate professor at the National Museum of Japanese History. His research interests are primarily Japanese ancient history and digital humanities research. He is currently engaged in a research project called “Integrated Studies of Cultural and Research Resources” and is conducting research on various techniques for digitizing historical data. He served as the chair of the Special Interest Group of Computers and Humanities under the auspices of the Information Processing Society of Japan from 2009 to 2010.

YUTA HASHIMOTO

Yuta Hashimoto, PhD, is an assistant professor at the National Museum of Japanese History, National Institutes for the Humanities. He is one of the co-directors of the largest crowdsourcing DH projects in Japan, *Minna-de-Honkoku*. He majored in the history of mathematics and mathematics education before entering the doctoral course and focusing his research on digital humanities. He has a bachelor’s degree from the Faculty of Literature but has also worked as an IT engineer, hence his interest in computer programming.

A. CHARLES MULLER

A. Charles Muller, PhD, is a professor at the University of Tokyo, in the Graduate School of Humanities and Sociology and its Center for Evolving Humanities. He specializes in Korean Buddhism and East Asian Yogācāra, and has published numerous books and articles on these topics. He is one of the earliest and most prolific developers of online research resources for the field of Buddhist Studies, being the founder and managing editor of the online *Digital Dictionary of Buddhism*, the CJKV-English Dictionary (<http://www.acmuller.net>), the *H-Buddhism Scholars Information Network*, along with having digitized and published numerous reference works.

MASAHIRO SHIMODA

Masahiro Shimoda, PhD, is a professor in the Department of Indian Philosophy and Buddhist Studies, within the Graduate School of Humanities and Sociology, and also the director of the Digital Humanities Initiative, within the Center for Evolving Humanities, at the University of Tokyo. He specializes in the history of the formation of the Buddhist scriptures, which elucidates the production and the process of passing down the

scriptures of traditional Indian Buddhism. As the head of the Center for Evolving Humanities, established in 2013, he teaches the “General Introduction for Digital Humanities” (<http://dh.iii.u-tokyo.ac.jp/>) as a core module of the center.