



Text and transcription

Sean M. Winslow

CLARIAH-AT
TEI Introductory School
10.-13. September 2019

Venue:

Centre for Information Modelling (ZIM)
Elisabethstraße 59/III, SR 81.31

Programm

Wednesday 11.9.

- 9:00-10:30 *Metadata: teiHeader, msDesc*
 Helmut W. Klug, Elisabeth Raunig
- 11:00-12:30 *Text and transcription*
 Sean M. Winslow
- 14:00-15:30 *Practice*
 Helmut W. Klug, Elisabeth Raunig, Sean M. Winslow
- 16:00-18:00 *Presentations and social event*

Document Structure

```
<TEI>
  <HEADER/>
  <TEXT>
    <BODY>
      <DIV>
        <HEADER/>
        <P/>
      </DIV>
    </BODY>
  </TEXT>
</TEI>
```

← You Are Here
←
←
←
←
←

Maximal Encoding Example

[Henry V. folgerdigitaltexts.org/download/](http://Henry.V.folgerdigitaltexts.org/download/)

TEI text body div1 div2 sp ab w

```
21276  </speaker>
21277  ▼ <ab xml:id="ab-1086">
21278  <lb xml:id="lb-10855"/>
21279  < milestone unit="ftln" xml:id="ftln-1086" n="3.1.1" ana="#verse" next="#ftln-1087" corresp="#w0171410
21280  <w xml:id="w0171410" n="3.1.1">Once</w>
21281  <c xml:id="c0171420" n="3.1.1"> </c>
21282  <w xml:id="w0171430" n="3.1.1">more</w>
21283  <c xml:id="c0171440" n="3.1.1"> </c>
21284  <w xml:id="w0171450" n="3.1.1">unto</w>
21285  <c xml:id="c0171460" n="3.1.1"> </c>
21286  <w xml:id="w0171470" n="3.1.1">the</w>
21287  <c xml:id="c0171480" n="3.1.1"> </c>
21288  <w xml:id="w0171490" n="3.1.1">breach</w>
21289  <pc xml:id="p0171500" n="3.1.1">, </pc>
21290  <c xml:id="c0171510" n="3.1.1"> </c>
21291  <w xml:id="w0171520" n="3.1.1">dear</w>
21292  <c xml:id="c0171530" n="3.1.1"> </c>
21293  <w xml:id="w0171540" n="3.1.1">friends</w>
21294  <pc xml:id="p0171550" n="3.1.1">, </pc>
21295  <c xml:id="c0171560" n="3.1.1"> </c>
21296  <w xml:id="w0171570" n="3.1.1">once</w>
21297  <lb xml:id="lb-10860"/>
21298  < milestone unit="ftln" xml:id="ftln-1087" n="3.1.2" ana="#verse" prev="#ftln-1086" rend="turnunder" co
21299  <w xml:id="w0171580" n="3.1.2">more</w>
```

Maximal Encoding Example

[Henry V. folgerdigitaltexts.org/download/](http://Henry.V.folgerdigitaltexts.org/download/)

```
74 ▼      <editorialDecl>
75 ►        <correction> [4 lines]
80 ►        <quotation> [5 lines]
86 ►        <hyphenation eol="hard"> [10 lines]
97 ▼        <segmentation>
98 ▼          <p>Acts are enclosed in <gi>div1</gi> tags. Scenes are enclosed in <gi>div2</gi> tags.
99            Prologues, epilogues, inductions, and choruses are tagged using either <gi>div1</gi> or
100           <gi>div2</gi>, whichever is most appropriate.</p>
101 ▼          <p>Page breaks from the print edition are included in <gi>pb</gi> tags. Line breaks are
102            enclosed in <gi>lb</gi> tags. The <att>edRef</att> attribute is used to identify line
103            breaks that differ between editions.</p>
104 ▼          <p>Speeches are enclosed in <gi>sp</gi> tags. Speech tags usually contain a speech prefix
105            (in the <gi>speaker</gi> tag), and an anonymous block (<gi>ab</gi>) containing the text
106            of the speech.</p>
107 ▼          <p>Stage Directions are enclosed within <gi>stage</gi> tags. They may occur within or
108            between speeches. Where appropriate, <att>who</att> attributes specify who participates
109            in the stage direction.</p>
110 ▼          <p>Poems, songs, and letters are enclosed in <gi>seg</gi> tags, with the <att>type</att>
111            attribute specifying the type. Segments split between speeches are linked through
112            <att>next</att> and <att>prev</att> attributes. Some songs are prefaced with a header
113            like "The Song", which is enclosed in a <gi>label</gi> tag.</p>
114          <p>Foreign words and phrases are enclosed in <gi>foreign</gi> tags.</p>
115 ▼          <p>Words, spaces, and punctuation marks are enclosed in <gi>w</gi>, <gi>c</gi>, and
116            <gi>pc</gi> tags respectively.</p>
117 ▼          <p>Lines (verse and prose) are not segmented within <gi>l</gi> or <gi>line</gi> elements.
118            They are noted through line breaks and <gi>milestone</gi> elements.</p>
119          </segmentation>
```

More Minimal Examples

tei-c.org/release/doc/tei-p5-doc/en/html/examples-editorialDecl.html

```
<editorialDecl>
<normalization>
  <p>All words converted to Modern American spelling using
      Websters 9th Collegiate dictionary
  </p>
</normalization>
<quotation marks="all">
  <p>All opening quotation marks converted to " all closing
      quotation marks converted to &#8226;</p>
</quotation>
</editorialDecl>
```

```
<editorialDecl>
<correction>
  <p>Turned letters are silently corrected.</p>
</correction>
<normalization>
  <p>Original spelling and typography is retained, except that long s and ligatured
      forms are not encoded.</p>
</normalization>
</editorialDecl>
```

Do What You Want, but Document it in EncodingDesc



Check the Documentation

The screenshot shows the TEI P5 Guidelines for Electronic Text Encoding and Interchange. The page title is "P5: Guidelines for Electronic Text Encoding and Interchange". The sub-page title is "<supplied>". The text defines "<supplied>" as text supplied by the transcriber or editor for any reason, such as damage or omission. It lists attributes like @xml:id, @n, @xml:lang, @xml:base, @xml:space, @att.global.rendition, @att.global.linking, @att.global.analytic, @att.global.facs, @att.global.change, @att.global.responsibility, @att.global.source, @att.editLike, @att.dimensions, @att.ranging, and @att.reason. The "Status" is optional, and the "Datatype" is 1-∞ occurrences of teidata.word separated by whitespace. Red arrows point from the right margin to the "Status" and "Datatype" definitions.

P5: Guidelines for Electronic Text Encoding and Interchange

Version 3.6.0. Last updated on 16th July 2019, revision daa3cc0b9

<supplied>

<supplied> signifies text supplied by the transcriber or editor for any reason; for example because the original cannot be read due to physical damage, or because of an obvious omission by the author or scribe. [11.3.3.1 Damage, Illegibility, and Supplied Text]

Module	transcr — Representation of Primary Sources
Attributes	<code>att.global (@xml:id, @n, @xml:lang, @xml:base, @xml:space) (att.global.rendition (@rend, @style, @rendition)) (att.global.linking (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) (att.global.analytic (@ana)) (att.global.facs (@facs)) (att.global.change (@change)) (att.global.responsibility (@cert, @resp)) (att.global.source (@source)) att.editLike (@evidence, @instant) att.dimensions (@unit, @quantity, @extent, @precision, @scope) (att.ranging (@atLeast, @atMost, @min, @max, @confidence))</code> @reason one or more words indicating why the text has had to be supplied, e.g. <i>overbinding, faded-ink, lost-folio, omitted-in-original.</i> Status Optional Datatype 1-∞ occurrences of <code>teidata.word</code> separated by whitespace

c. 11: tei-c.org/release/doc/tei-p5-doc/en/html/PH.html

 <Text Encoding Initiative>

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines — English

P5: Guidelines for Electronic Text Encoding and Interchange

Version 3.6.0. Last updated on 16th July 2019, revision daa3cc0b9

<p>Table of contents</p> <ul style="list-style-type: none">11.1 Digital Facsimiles11.2 Combining Transcription with Facsimile11.3 Scope of Transcriptions11.4 Advanced Uses of surface and zone11.5 Aspects of Layout11.6 Headers, Footers, and Similar Matter11.7 Identifying Changes and Revisions11.8 Other Primary Source Features not Covered in these Guidelines11.9 Module for Transcription of Primary Sources <hr/> <p>» 10 Manuscript Description » 12 Critical Apparatus Home</p>	<h2>11 Representation of Primary Sources</h2> <p>This chapter defines a module intended for use in the representation of primary sources, such as manuscripts or other written materials. Section 11.1 Digital Facsimiles provides elements for handling digitally-encoded images of such materials. This module may also be useful in the preparation of critical editions, but the module defined here is distinct from that defined in chapter 12 Critical Apparatus, and may be used independently of it. Detailed metadata relating to primary sources of any kind may be recorded using the elements defined by the manuscript description module discussed in chapter 10 Manuscript Description, but again the present module may be used independently if such data is not required.</p> <p>Although this chapter discusses manuscript materials more frequently than other forms of written text, most of the recommendations presented are equally applicable <i>mutatis mutandis</i> to the encoding of printed matter or indeed any form of written source, including monumental inscriptions. Similarly, where in the following descriptions terms such as 'scribe', 'author', 'editor', 'annotator' or 'corrector' are used, these may be re-interpreted in terms more appropriate to the medium being transcribed. In printed material, for example, the 'compositor' plays a role analogous to the 'scribe', while in an authorial manuscript, the author and the scribe are the same person.</p> <h3>1 11.1 Digital Facsimiles</h3> <p>These Guidelines are mostly concerned with the preparation of digital texts in which pre-existing sources are transcribed or otherwise converted into character form, and marked up in XML. However, it is also very common practice to make a different form of 'digital text', which is instead composed of digital images of the original source, typically one per page, or other written surface. We call such a resource a <i>digital facsimile</i>. A digital facsimile may, in the simplest case, just consist of a collection of images, with some metadata to identify them and the source materials portrayed. It may sometimes contain a variety of images of the same source pages, perhaps of different resolutions, or of different kinds. Such a collection may form part of any kind of document, for example a commentary of a</p> <p>» 11.2 Combining Transcription with Facsimile Home</p>
---	--

Example: Wellcome Library, Misc. Cookery Books

wellcomelibrary.org/item/b19030253#?c=0&m=0&s=0&cv=48&z=-0.3899%2C-0.1193%2C1.8624%2C1.1699

To make a calefs foot pye

24

Bake y^r calefs foot after they are scalded and boyled and put ym
into a good crust with a little butter and rosemary and currans and bake it so
and when it is enough cutt of ye litle and put into it a glass of whitem wine
mixt with som butter and suger and a little vinger and ye yolk of an egg
and made very hott over ye fire.

To make y^r ~~good~~^{sweet} oynment a most excellent and
approved good one / for paine in ye lins or a cruch

Bake sage ~~one~~ one pound bay leaves and wormwood one pound melleket
herbes and flowers of canamile spicke roesmary Led roses ¹/ Johnwort and
march mallows of each a good handfull chop all ye herbes as small as can be

PB (Page Beginning)

To make a calefs foot pye

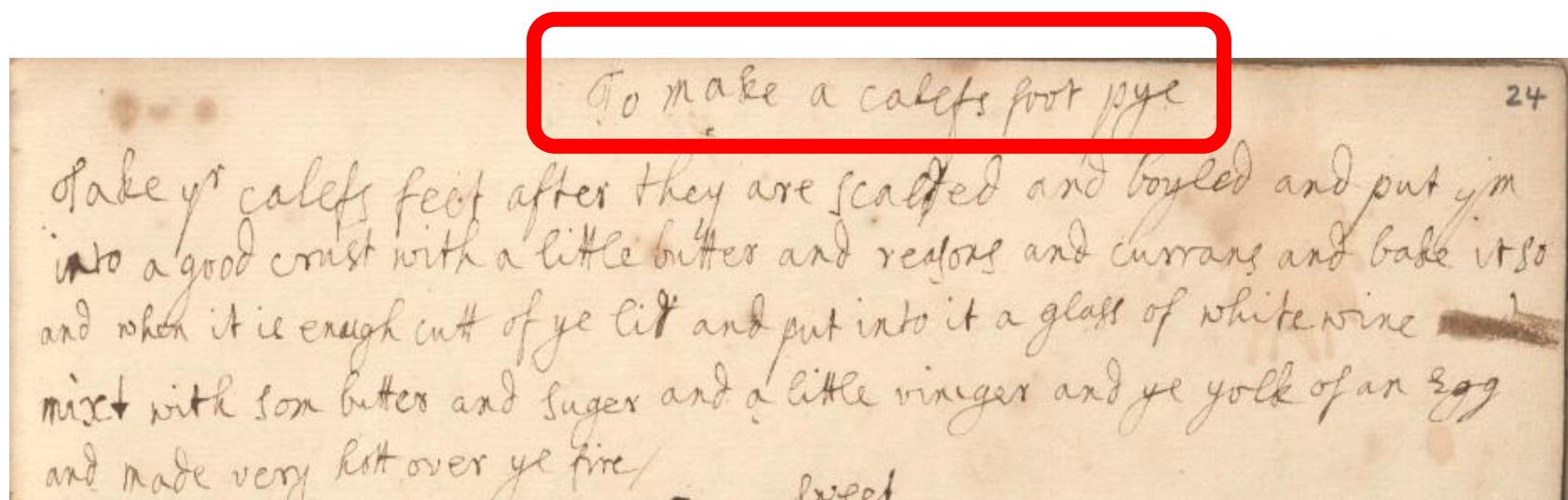
24

Take y^r calefs foot after they are scalded and boyled and put y^m into a good crust with a little butter and reasone and currans and bake it so and when it is enough cutt of ye litle and put into it a glass of white wine
mixt with som butter and suger and a little vinger and ye yolk of an egg
and made very hott over ye fire

Sweet

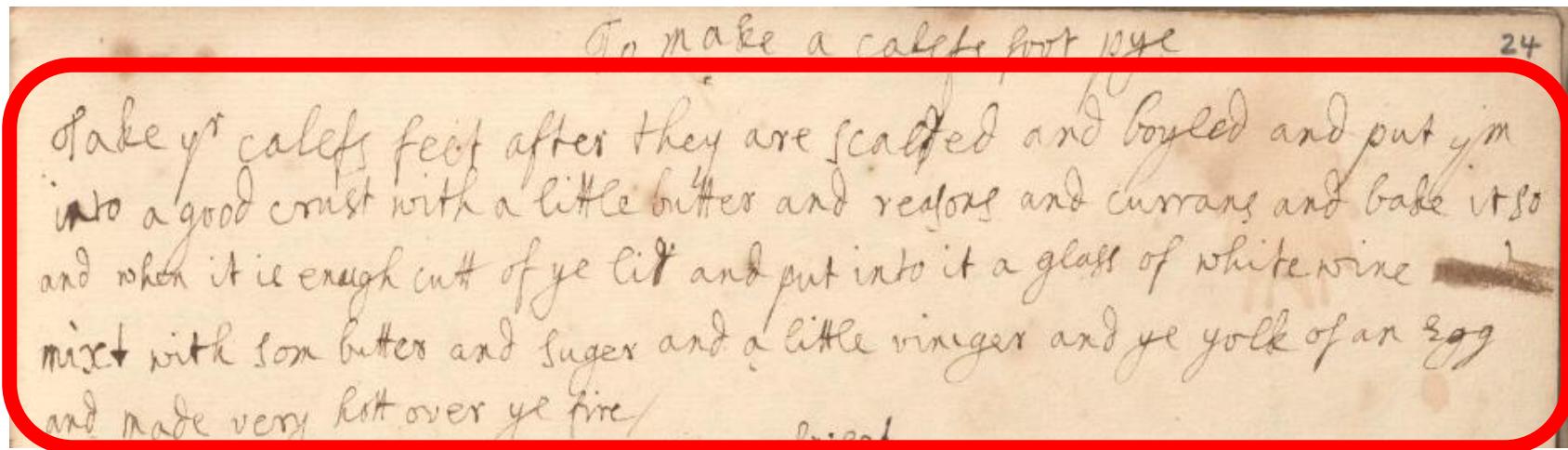
<pb n=24>

HEAD (Any-Level Header)



```
<div>
  <head>To make a calefs foot
  pye</head>
```

P (Paragraph)



< p > Take yr calefs feet after they are
scalded and boyled and put ym into
a good crust ... </ p >

(Some) Transcription Elements

<unclear>

<gap>

<damage>

<supplied>

<sic>

<add>

<subst>

<restore>

<retrace>

<delSpan>

<addSpan>

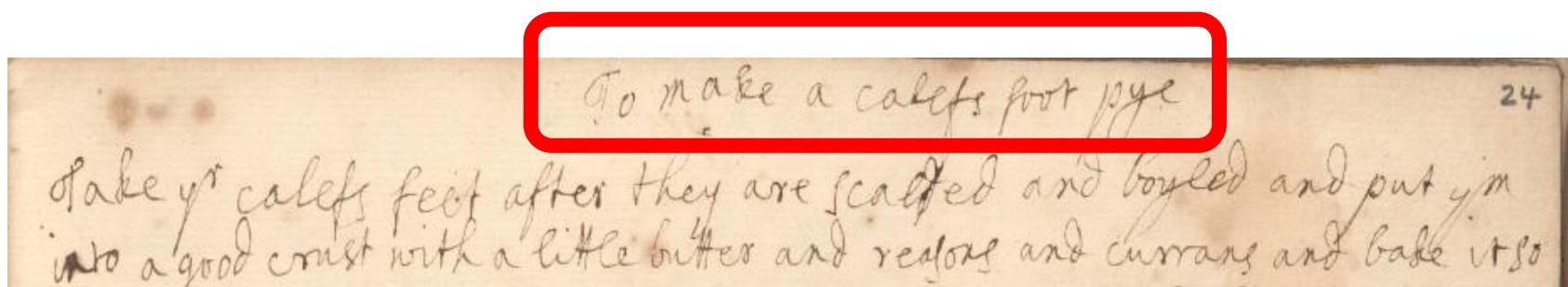
<damageSpan>

<choice>

<orig> / <reg>

<abbr> / <expan>

SIC (*Sicut*, “Just So”)



<head>

To make a

<sic>calefs</sic> foot

<sic>pye</sic>

</head>

CHOICE / ORIG / REG

To make a calefs foot pye 24
Take y^r calefs foot after they are scalded and boyled and put y^m into a good crust with a little butter and reasone and currans and bake it so

<choice>

<orig>calefs</orig>

<reg>calf's</reg>

</choice> foot ...

CHOICE / ABBR / EXPAN

To make a calefs foot pye 24
Take yr calefs feet after they are scalded and boiled and put ym
into a yd crust with a little butter and reasone and currans and bake it so

<p>Take

<choice>

<abbr>yr</abbr>

<expan>your</expan>

</choice> calefs feet after they ...

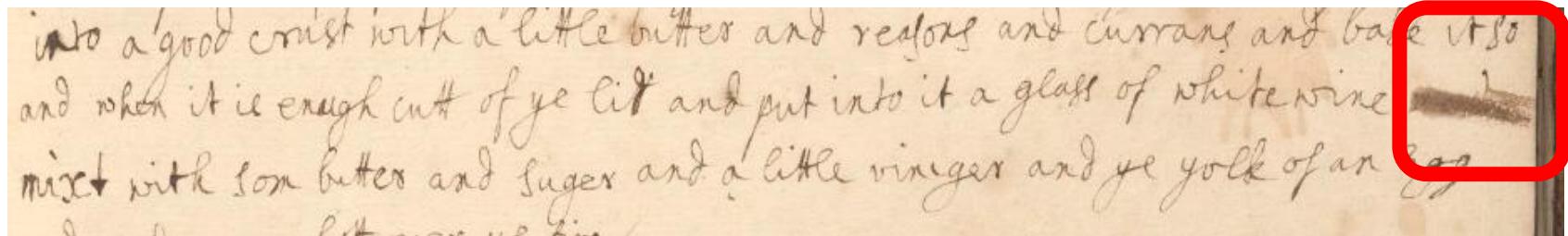
UNCLEAR

To make a calefs foot pye 24
Take yr calefs feet after they are scalded and boyled and put ym
into a good crust with a little butter and reyons and currans and bake it so

... after they are

<unclear reason="illegible">scalded
</unclear> and boyled ...

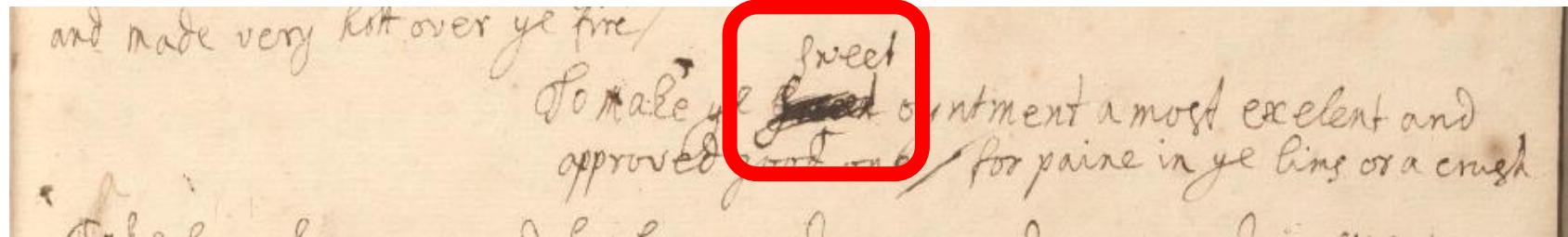
GAP (cannot be read)



into a good crust with a little butter and raisons and currans and bake it so
and when it is enough cut of ye lift and put into it a glass of white wine
mixt with som butter and sugar and a little vinger and ye yolk of an egg

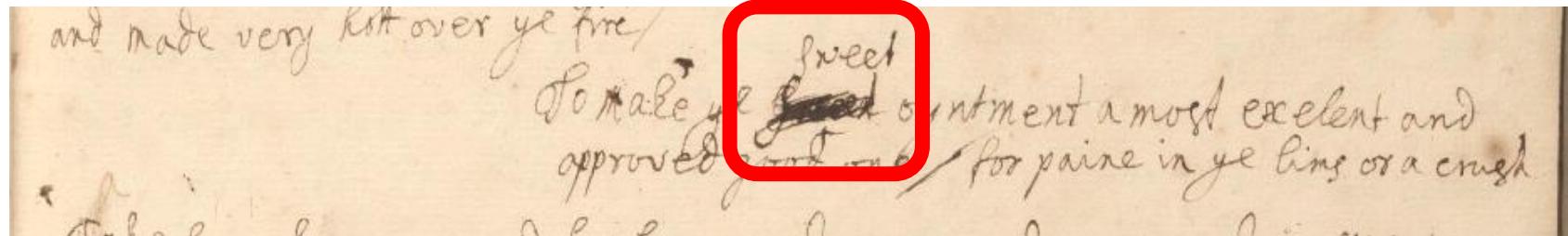
... put it into a glass of white wine
<gap reason="smudge" extent="1 word"/> mix ...

DEL (Deletion)



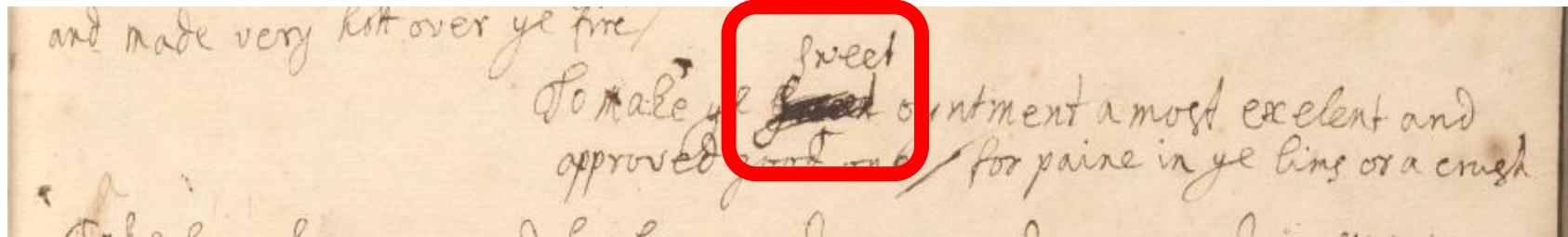
```
<del>  
    s<gap reason="cancelled"  
        quantity="3" unit="chars"/>t  
</del>
```

ADD (Addition)



<head>To make ye
<add place= "above">
 sweet
</add>
oyntment a most exelent ...

SUBST (Substitution)



<subst>

```
<del>s<gap reason="cancelled"  
quantity="3" unit="chars"/>t </del>  
<add place="above">sweet</add>  
</subst>
```

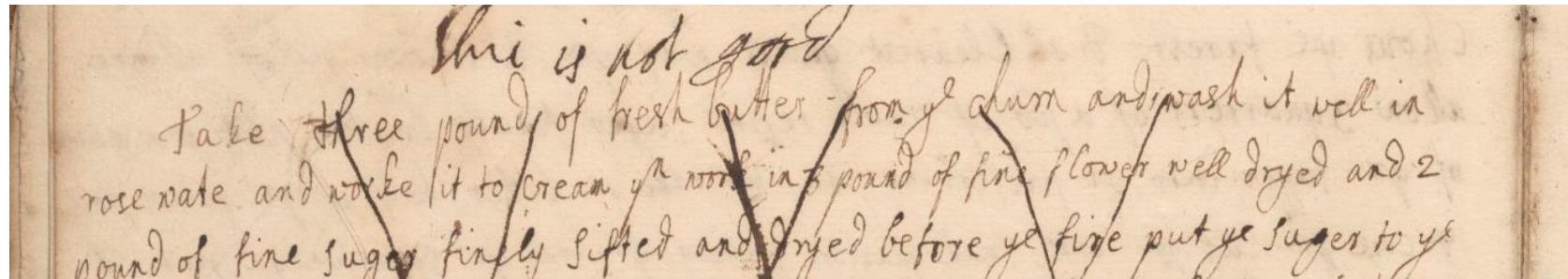
Example: Wellcome Library, Misc. Cookery Books

wellcomelibrary.org/item/b19030253#?c=0&m=0&s=0&cv=15&z=-0.1348%2C-0.0743%2C1.3198%2C0.829

To make a cake m^{rs} m^{rs} ^{hong}
This is not good

Take three pounds of fresh butter from ye alum and wash it well in
rose water and worke it to cream y^e work in 3 pound of fine flower well dried and 2
pound of fine sugar finely sifted and dryed before ye fire put ye sugar to ye
butter first y^e beat 30 eggs half ye whites taken out and put half a pint of
sack to y^e beat y^e very well y^e put to ye butter sugar and flower and put in
an ounce of mace finely beaten and 3 nutmeg and a little musk and amber
green striped in orangeflower water worke it very well and put into it a pd
of Jordan almonds either beaten with orangeflower water or sliced very
thinn leanthways and when you are going to put it into ye oven stroye in
6 pound of cynamon after they are washed and well dryed before ye fire 3 hours
will bake it

DELSPAN (Delete Span)



<add>This is not good</add>

<delSpan spanTo="#end_anchor"/> referenced with "#" closed

<p> Take three pounds of fresh ... </p>

<anchor xml:id="end_anchor"/> closed Unique XML:ID

Similar Span Elements

<delSpan> for

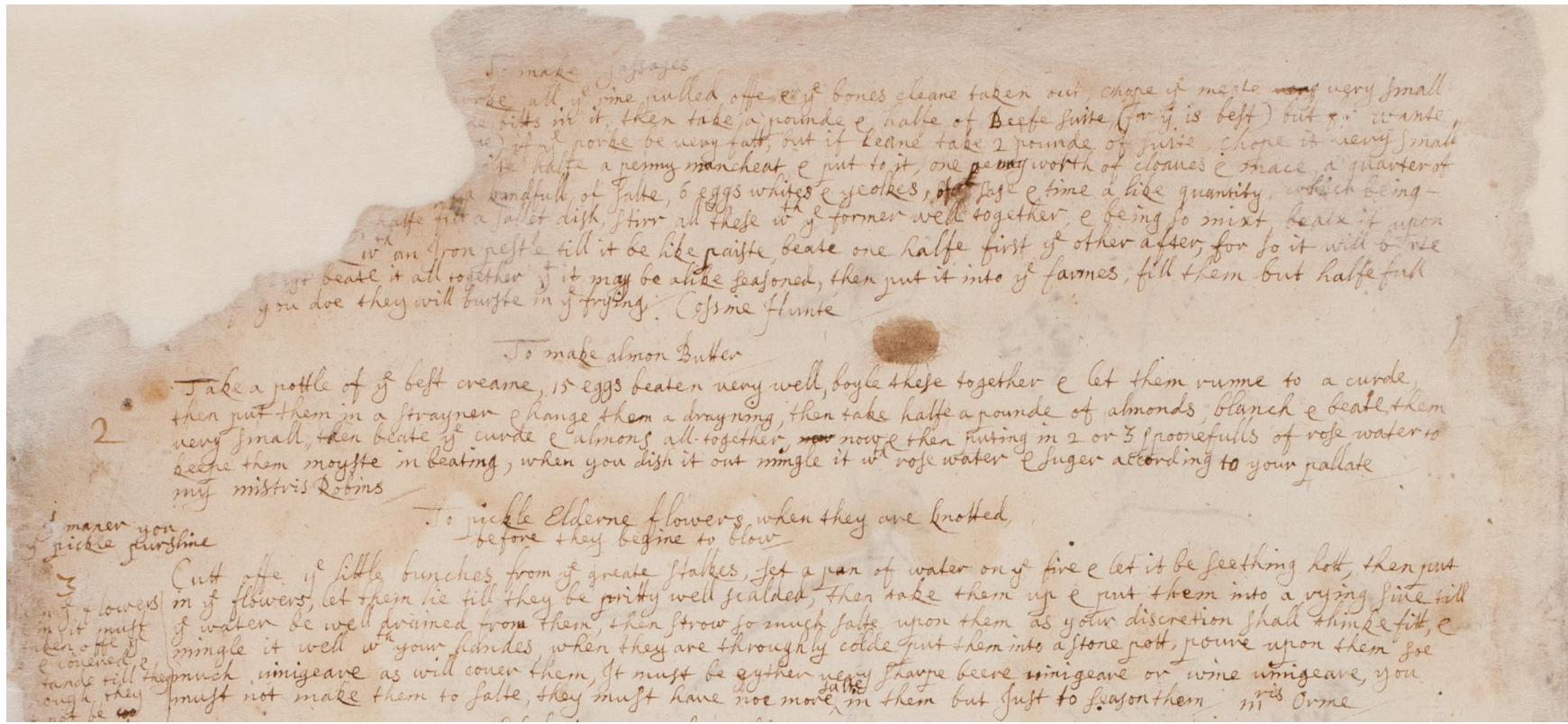
<addSpan> for <add>

<damageSpan> for <damage>

Example:

Anna Cromwell English Recipe Book, 1650, with Later American Additions (Charles Brigham Account Book)

<https://gigi.mwa.org/netpub/server.np?quickfind=271275&site=manuscripts&catalog=catalog&sorton=filename&template=detail.np&offset=8&TabletNPResults=/netpub/server.np%3Fquickfind%3D271275%26site%3Dmanuscripts%26catalog%3Dcatalog%26sorton%3Dfilename%26template%3Dresults.np&TabletNPResultsCount=406&playMode=stop>



DAMAGE (Damage)

To make sassages
orbe all ye rine pulled offe & ye bones cleane taken out, chose of meatte ~~is~~ very small
bits in it, then take a pounde & halfe of Beefe Suet (fat is best) but fat werte,
it not norke be very fat, but if Leane take 2 pounde of suet, chose it very smale
take a poynt maner & pur rope, one penyworth of cloaves & mace, a quarter of
a handfull of salt, 6 eggs whipes & yecokes, of fat & time a like quantity, which being
the hit a tasset dish, sturr all these wth former well together, & being so mixt, beat it upon

<head>To make sassages</head>

<damage agent="broken page">

<gap extent="about 1.5cm"/>

</damage>orbe all ye rine pulled
offe & ye bones cleane taken ...

Pair Exercise (Winslow Family Recipe Cards—Edited)

○ Russian Tea (Butter) Balls ○

1 Cup butter

2 Cups flour

4 Tblsp powdered sugar

1 Cup ~~nuts~~ haselnuts

1 tsp vanilla

Cream butter, ~~sugar~~, & vanilla, then add
flour & nuts. Bake 15-18 min at 350°.

Roll in powdered sugar while warm & again
after cool. ~ 3 dozen

LIST / ITEM & SUPPLIED

1 Cup butter

2 Cups flour

4 Tblsp powdered sugar

1 Cup ~~nuts~~ hazelnuts

1 tsp vanilla

<list>

<item>1 Cup butter</item>

Cream butter, ~~sugar~~, & vanilla, then add

Cream butter, <supplied
reason=>sugar,</supplied> & vanilla

Pair Exercise (Winslow Family Recipe Cards—Edited)

○ Russian Tea (Butter) Balls ○

1 Cup butter

2 Cups flour

4 Tblsp powdered sugar

1 Cup ~~nuts~~ haselnuts

1 tsp vanilla

Cream butter, ~~sugar~~, & vanilla, then add
flour & nuts. Bake 15-18 min at 350°.

Roll in powdered sugar while warm & again
after cool. ~ 3 dozen

Transcription Elements to Try

<unclear>

<gap>

<damage>

<supplied>

<sic>

<add>

<subst>

<restore>

<retrace>

<choice>

<orig> / <reg>

<abbr> / <expan>

<list>

<item>

<delSpan>

<addSpan>

<damageSpan>

Closing Thoughts

- These slides are just to give you an idea of the kinds of transcription elements available to you.
- None of this is a replacement for reading the TEI documentation.
- Choose a tag-set that works for your project, be consistent, and document it in the header.