

Newton Project: Transcription and Tagging Guidelines and XML Tag Set

(Version 6 for use with the Newton Project RNG schema)

John T. Young

with a lot of help from Michael Hawkins

and thanks to Linda Cross, Raquel Delgado-Moreira and Yvonne Martin-Portugues
for proofreading and constructive criticism

updated 3 July 2011

Introduction

These guidelines should be read in conjunction with the accompanying Element Set. The 'Document Structure' and 'Text' sections provide a transcription policy and a broad outline of which tags to use where, while the Element Set gives specific details of what exactly is required or permitted in each element and where it may (or may not) be used. It is (or at least aims to be) a plain English version of the Newton Project schema.

I've done my best to keep jargon to a minimum, but have assumed familiarity with certain key XML terms and the means of representing them: principally *element* (and the distinction between *empty* and *non-empty* elements), *entity*, *attribute* and *attribute value*. It is essential that transcribers are entirely clear about these terms, and understand the principle of *nesting* elements. These terms are not difficult to grasp and are explained in any guide to XML - though a good place to start if you're new to the language is the 'Gentle Introduction to XML' on the TEI (Text Encoding Initiative) website at <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>.

Throughout this document, I use the term '**text string**' to mean 'any quantity of continuous text': this may be a single letter within a word, a whole word, a sentence, five-and-a-half words, ten paragraphs or whatever. Element tags are referred to with their angle brackets on and empty element tags with their forward slash (e.g. `<pb/>`). Non-empty elements are referred to by their opening tag: '`<p>`', for instance, means the entire element `<p></p>` except where I specifically refer to 'the opening `<p>` tag'.

Please pay close attention to the rules for spacing before and after tags, particularly in regard to `<add>`, `` and `<lb />`. These may seem pettifogging but they are there for a reason. Incorrect use of spacing will lead to serious problems with the display of your transcript.

In most cases, the instructions apply equally to the transcription and encoding of manuscript and of print material, though obviously a number of tags, such as those for deleted and inserted text, are relevant only to transcription from manuscript, since these features do not normally occur in printed texts. There are, however, a few cases where the methodology for transcription from print is different from that for transcription from manuscript: this applies mainly to the elements `<anchor/>`, `<fw>`, `<note>`, `<pb/>` and `<teiHeader>`.

There are also a few distinctions to be drawn between the encoding of early modern (17th-18th century) and modern (19th century or later) printed documents. Transcribers will not normally be asked to key in or encode modern printed documents, as draft XML versions can be automatically generated from OCR scans, but they may be called on to proofread such drafts. Preliminary XML drafts of pre-nineteenth-century printed texts may also be automatically generated, depending on the quality of the images we have available. In either case, it is clearly essential for transcribers and proofreaders to be aware of the variant protocols for encoding such texts. These variants are clearly flagged at the relevant points in the guidelines.

Almost all the coding now used in Newton Project markup is derived from the TEI Guidelines (version P5), but we have made certain minor and TEI-compliant restrictions and extensions to suit the particular nature of the material we are dealing with.

Document Structure and the <teiHeader>

Each document is entirely enclosed in a so-called 'parent element', <TEI>. This is divided into two main component elements, <teiHeader> and <text> (always in that order). Another element, <body>, nests directly inside <text> (in theory, <text> can contain other things too, but these are not normally relevant to us). The transcription proper is wholly enclosed in the <body> element, which contains one or more <div>s, i.e. chapters, sub-chapters or other clearly defined structural divisions. The <teiHeader> is reserved for metadata: information about the source document, a record of the work that has been done on it, and the id values of the languages and hands that feature in it and of the transcribers and checkers who have worked on it. Thus every document has this outline:

```
<TEI>
  <teiHeader></teiHeader>
  <text>
    <body>
      <div></div>
      <div></div>
      <div></div>
    <!--there may be any number of <div>s bar zero;
    there may be <div>s within <div>s if necessary-->
    </body>
  </text>
</TEI>
```

Unless given special instruction, transcribers do not need to concern themselves about most aspects of the <teiHeader>. You will be given a template header to get you started and the editors will modify the fields as necessary after you've completed your transcript. The only part of the <teiHeader> that we do need you to fill in is the final section, <revisionDesc>, which contains a series of <change> elements recording the dates between which various stages of the transcription process took place. It follows this format:

```
<revisionDesc>
  <change when="2008-03-24"><name>John Young</name>
  began tagged transcription</change>
  <change when="2008-04-19"><name>John Young</name>
  completed tagged transcription</change>
  <change when="2009-04-20"><name>Michael
  Hawkins</name> began checking against
  microfilm</change>
  <change when="2009-05-01"><name>Michael
  Hawkins</name> finished checking</change>
</revisionDesc>
```

Simply enter the relevant dates as the when values of <change>, using the yyyy-mm-dd format as above, your name as the content of <name> and a brief account of the work you've begun or completed as the content of <change>.

NB: Keeping an exact account of the dates on which the various stages of processing a document are started and completed is very important because each upgrade of the tagging policy means that previous mark-up also has to be upgraded to bring it into line with the new dispensation. This process can be largely automated, but *only* if we have a record of what policy was being followed in the first place. If this isn't clear, the whole thing has to be proofread again from scratch. It also provides an easy way of verifying which is the latest version of a given document if any confusion arises in the back-up system.

The Text

a) General

1) The golden rule is: '**If in doubt, say so**'. Admitting to being confused or uncertain at any point is no shame at all. Being confused or uncertain and *not* admitting it merits boiling in oil. Please don't hesitate to contact an editor if at all unsure what to do at any point, however trivial the issue may seem.

2) All such doubts, if not resolved by personal communication, should be expressed in a `<!-- -->` tag, which can also be used for any **comments or queries** you want to make about the text. The comment appears between the two double-dashes, and may contain pretty well anything apart from another double-dash (single dashes are OK). It should be preceded by a capitalised category indicator from the following list:

```
<!-- TODO JY/MJH/NP -->: problem requiring action by John
Young/Mike Hawkins/unspecified Newton Project member
<!-- TRANSC -->: to comment on any difficulties or oddities associated
with the transcription or tagging
<!-- CODIC -->: to comment on codicological features of the
manuscript that are not covered by the Guidelines as they stand
<!-- APP -->: comments that may be useful to the editors when
producing an editorial apparatus
<!-- OTHER -->: anything else
```

See the Element Set for examples of how to use these. Transcribers are strongly urged to make liberal use of this tag.

You can put a `<!-- -->` tag anywhere at all in the text, but *not* inside other tags: e.g.

```
<add place="marginRight" indicator="no"><!--TRANSC
well, I think this is an <add> but perhaps we
should call it <note> - jy-->according to
Ierome</add>
```

is fine; so is

```
<add place="marginRight" indicator="no">according
to Ierome</add><!-- TRANSC well, I think this is an
<add> but perhaps we should call it <note> - jy-->
```

but *not*

```
<add <!-- TRANSC well, I think this is an <add> but
perhaps we should call it <note> - jy-->
place="marginRight" indicator="no">according to
Ierome</add>.
```

Note that the `<!-- -->` tag should come slap bang next to whatever it comments on, with no whitespace in between. Think of it as invisible (because it will be to the user), so the spacing should be exactly as it would be if it weren't there.

3) **Punctuation and spelling** (including mistakes) should be transcribed exactly as they appear in the original. Obvious **errors**, however, should be placed in a `<choice>` tag with the faulty version in `<sic>` and the corrected version in `<corr>`, e.g.

`'<choice><sic>superstion</sic><corr>superstition</corr></choice>'`. But there is no need to tag or comment on perfectly normal 17/18-century spellings such as 'beleive', 'recon', 'apostacy', or standard early modern Latin that looks 'wrong' to a classicist, such as 'authoritas' (which a classicist would spell 'auctoritas'). If uncertain as to whether an odd-looking spelling is a mistake or just an old form, say so in a `<!--TRANSC -->` tag.

4) **Capitalisation** should follow the original as far as possible, but there are times when it's very hard to say whether or not an initial letter *is* capitalised. Letters such as C and S, which don't necessarily much change their form between capital and lower-case, are particularly tricky. My suspicion is that half the time Newton himself wouldn't have known or cared which case he was using. This is one instance where we can to some extent be guided by the sense - all other things being equal, 'Christ' and 'cow' seem likelier than 'christ' and 'Cow', and 'Spiritus Sanctus' and 'seven' likelier than 'spiritus sanctus' and 'Seven'. But this has to be to some extent a subjective decision. Again, you can hedge your bets by using a `<!--TRANSC -->` tag.

Transcribe initial 'ff' (which functions as a modern capital 'F') as the element `&ff;`.

Neither Newton nor the majority of his contemporaries distinguish in manuscript between capital I and J, or between capital U and V, so as a general rule, transcribe 'Iesus', 'Iewes', 'Iames', 'Vnlesse', 'Vnction', etc.¹ But if you are dealing with a printed work or you come across a hand that does make a distinction, obviously you should follow suit, so long as you're sure the distinction really is there on the page and not just in your mind. Newton does, however, distinguish between lower-case i and j, and between u and v, and again we should follow him rather than modernise/standardise in instances such as 'Petavij', 'ijsdem', or the use of 'j' as the Roman numeral for '1'.

5) Record the use of **brevigraphs** (i.e. conventional symbols denoting a particular letter or series of letters, such as overlining, tails meaning 'us' or 'ue' in Latin passages, crossed 'p's meaning 'per', 'par', 'pre' or 'pro' and the like) by using the `<orig>` and `<reg>` tags as explained in the Element Set.

6) **Other conventional abbreviations** (in Newton's case these nearly always involve the use of superscripted letters) should be transcribed exactly as they appear in the original. Place the abbreviated form in `<abbr>` and the expanded version in `<expan>`, and nest both in a `<choice>` element, e.g. `'<choice><abbr>y<hi`

¹ In Keynes Ms. 3, p. 21, Newton alters the word 'Iews' to read 'Israel', which he does by writing 'srael' over 'ews': the initial letter is untouched. In a quasi-alphabetical list of terms in the Pierpont Morgan notebook, f. 42v, the words 'Victualling-house' and 'Vtensills' are both listed under V.

rend="superscript">r</hi></abbr><expan>your</expan></choi
 ce>', '<choice><abbr>S<hi
 rend="superscript">r</hi></abbr><expan>Sir</expan></choic
 e>', '<choice><abbr>B<hi
 rend="superscript">p</hi></abbr><expan>Bishop</expan></ch
 oice>' for 'y^r', 'S^r', 'B^p'. Common abbreviations such as þ <hi
 rend="superscript">e</hi> for 'the', w<hi
 rend="superscript">ch</hi> for 'which', etc. are not included in the Entity
 Set as not all scribes abbreviate them in exactly the same way. File-specific entity
 declarations for &the;, &that;, &with;, &which;, etc., or any other abbreviations that
 occur frequently in a given file, may be added by transcribers (guidance will be
 offered if necessary on how to declare these). Conventional abbreviations that are
 still standard, such as 'Dr', 'Mr', 'A.D.' - or, in references, 'b.', 'c.' and 'v.' for 'book',
 'chapter' and 'verse' - do not need to be tagged, but all others should be.

Transcribers should *not* attempt to expand curtailed forms of proper nouns, or
 any idiosyncratic abbreviations, as these are a minefield. It is often by no means
 obvious whether 'Ier.', for instance, means 'Ierome', 'Ieremiah' or 'Ierusalem'. This is a
 job for the academic editors - though as ever, comment tags are welcome.

7) **Catchwords, page numbers, running headers, shelfmarks and sigils** are dealt
 with using <fw> ('**forme work**'). Page numbers in manuscripts or early printed texts
 should appear immediately after the page break (<pb/>) in the transcript, irrespective
 of where they actually appear on the page (this is indicated by the place value).

No space should be left between </fw> and <pb/> or between <pb/> and
 <fw>: for instance 'without noting any various lections in <fw
 type="catch" place="bottomRight">them</fw><pb
 xml:id="p022r" n="22r"/><fw type="pag"
 place="topRight">22</fw> them'.

If a catchword is incomplete, no space should be left between the component
 parts of the word. Otherwise, leave one space either side of the
 <fw></fw><pb/><fw></fw> sequence. For instance:

'I <fw type="catch">under</fw><pb/><fw
 type="pag">17</fw>understand'

'I under<fw type="catch">stand</fw><pb/><fw
 type="pag">17</fw>stand'

'I <fw type="catch">understand</fw><pb/><fw
 type="pag">17</fw> understand'.

If a catchword is incomplete and it and/or the preceding word has a **hyphen**,
 indicate this with <lb type="hyphenated"/>. For instance:

'I <fw type="catch">under<lb
 type="hyphenated"/></fw><pb/><fw
 type="pag">17</fw>understand'

'I under<lb type="hyphenated"/><fw
 type="catch">stand</fw><pb/><fw type="pag">17</fw>stand'.
 <lb/> is not otherwise needed before, after or within <fw>.

Be aware of the distinction between **page break** (see section b) 5 below) and
page number. Page break (<pb/>) indicates the physical point at which the text
 moves on to a new page; page number (<fw type="pag">) encodes page numbers
 that actually feature in the document (whether they were put there by the original

writer or anyone else, and whether or not they correspond to the number assigned to that page by the transcriber/encoder).

Variant procedure for modern printed texts

In transcriptions of modern printed texts (nineteenth-century or later), `<fw>` is not needed at all (since there are normally no catchwords in modern texts and the page number will always be the same as the `n` value of `<pb/>`). However, transcriptions of seventeenth- or eighteenth-century printed texts should follow the protocol for manuscripts, as these often feature catchwords, and mispagination is quite common: this can be indicated by using `<sic>` and `<corr>`. For example, if p. 29 has been mispaginated as p. 92, encode it as '`<pb xml:id="p029" n="29"/><fw type="pag" place="topRight"><choice><sic>92</sic><corr>29</corr></choice></fw>`'.

8) **Uncertain or conjectural readings** should be tagged `<unclear>`, with the degree of certainty expressed in the `cert` value on a scale of "high" (pretty confident), "medium" (doubtful) or "low" (an educated guess). The reason for the uncertainty should be stated by the `reason` value (see Element Set for permitted values).

Where two or more readings are plausible, group them as separate `<unclear>` elements nesting in `<choice>`. The `cert` values can be used to weight the relative plausibility of the alternatives: e.g. '`<choice><unclear cert="high">goal</unclear><unclear cert="low">goat</unclear></choice>`' means 'I'm pretty sure this says "goal" but it might just say "goat"', while '`<choice><unclear cert="medium">goal</unclear><unclear cert="medium">goat</unclear></choice>`' means 'I think this says either "goal" or "goat" but I don't know which'.

9) Words or passages that are **missing** or are wholly **illegible** for whatever reason obviously can't be transcribed. Indicate the omission with a `<gap/>` tag, using the `unit`, `extent` and `reason` values to explain how much text is omitted from the transcription (if this can be ascertained) and why. If it's possible to make a reasonable guess as to what the missing or illegible text should be, use `<supplied>` instead of `<gap/>` (see section a) 10 below).

10) Material that is wholly **illegible**, has been **omitted**, or has completely **disappeared** from the document (usually through manuscript damage), but is at least conjecturally recoverable from the context or by reference to another version of the same text, should be tagged `<supplied>`. In the latter case, the source, if it is not just the transcriber's common sense, should also be recorded as the `source` value.

11) Words or passages in **a language different from that of the surrounding text** should be tagged `<foreign>` with `xml:lang` values as declared in the `<langUsage>` section of the `<teiHeader>`. If such passages violate element boundaries, use successive nested `<foreign>` tags. `<foreign>` may nest within `<foreign>`, so if for instance there's a bit of Greek in the middle of a Latin passage in a document whose main language is English, it is tagged thus:

```
main English text of document <foreign
xml:lang="lat">Latin interpolation with a
<foreign xml:lang="gre">bit of Greek</foreign>
in the middle of it</foreign> resumption of
English text
```

Greek and Hebrew characters, Latin characters with diacritics (such as accents and cedillas), and Latin ligatures (e.g. æ, œ) should be encoded using the entities defined in the Entity Set. Greek ligatures should be silently expanded, if you know how to decipher early modern Greek ligatures. Greek and Hebrew passages, however, should only be attempted by transcribers reasonably confident of their ability to read them; otherwise just put in a 'todo' tag, e.g. '`<!--TODO NP I don't know how to read this ligature-->`'; '`<!--TODO NP five words in Hebrew-->`'. If you *can* transcribe Hebrew, do it in sense-order. Though the Hebrew characters and words read from right to left in the original, they should be transcribed left to right, i.e. the rightmost character in the original becomes the leftmost entity in the transcription. (It comes out the right way round in the browser.) If confronted with Arabic (or any other script not covered by the Entity Set), speak to an editor.

b) Tagging for layout and spacing

Note that *all* formatting, be it layout, character style, font size or whatever, has to be indicated by tags. So does **whitespace** (i.e. gaps of more than one space or line between individual words, characters or lines).

1) **Headings**, whether of the whole document or of a section within it, should be tagged `<head>`. The `rend` value indicates whether or not they are centered. Headings are usually in a somewhat larger script than the body text, but there is no need to record this in the tags. The tag `<head>` should only occur at the beginning of a `<div>` or `<lg>` (line group, i.e. verse passage). **Things that look like headings but do not in fact introduce new sections** should be tagged `<floatingHead>` (see the Element Set for details).

2) **Paragraphs** should each be enclosed in a `<p>` tag: irregular indentation of the first line is indicated by the `rend` value. No `rend` value is needed if the indentation is normal (i.e. the first line of the paragraph is indented by about 3-5 spaces). Don't bother recording slight variants in indentation, only ones pronounced enough to seem potentially significant, e.g. not indented at all or indented by about ten or fifteen spaces. If your document or one of its `<div>`s (or one of its pages) begins or ends in the middle of a paragraph, the rules of markup language dictate that you still have to tag it as though the paragraph were complete - point out the anomaly in a `<!-- TRANSC -->` tag.

If an entire paragraph is right-aligned (this is very rare), tag it as `<p rend="right">`. If you encounter weirder indentation such as 'hanging indents' (where the first line of the paragraph is not indented at all but all the subsequent ones are), simply note the issue in a `<!-- TRANSC -->` tag.

3) **Line breaks in prose** should be indicated by `<lb/>`. If the line break occurs between words, place it immediately before the second word, leaving one space after the first: 'the Visigoths reigned but `<lb/>`three years'. If it occurs in the middle of a word, leave no spaces: 'founded the king`<lb/>`dom of the Franks'. Don't record line breaks in interlinear and marginal insertions or notes.

If a word is **hyphenated** because it has a `<lb/>` in the middle of it, *don't* transcribe the hyphen (unless the hyphen is part of a compound word and would have appeared anyway, as in 'Idol-`<lb/>`Temples' - if in doubt, put in a `<!-- TRANSC -->` tag). Instead, apply the attribute `type` with the value "hyphenated", e.g. 'Idol`<lb type="hyphenated"/>`atrous'.

`<lb/>` is not required at the beginning or end of a page, paragraph or line of verse, or before or after a catchword unless the catchword (or the preceding text string) is a partial word and hyphenated, in which case use `<lb type="hyphenated"/>`.

Line breaks that appear to have been introduced for a purpose (usually in headings where the layout can be construed as having some semantic significance) should be tagged `<lb type="intentional"/>` (see under `<lb/>` in the Element Set for an example).

4) **Verse passages** (even if they are only one line long) should be tagged `<lg>` (line group) instead of `<p>`. The individual lines should each be enclosed in `<l>` (line) and do not need to be introduced by `<lb/>`. Irregular indentation is indicated by the `rend` value of `<l>` (as with `<p>`). If the poem has a **heading**, the whole thing (including the heading, tagged `<head>`) needs to be tagged as `<lg>`: this can have smaller line groups (stanzas) within it if necessary. Spacing around `<l>` is irrelevant for processing purposes, but for ease of proofing the best thing to do is to start each `<l>` on a new line of your transcript.

5) **Page breaks** (including the beginning of the first page of a document) should be indicated by `<pb/>`. In manuscripts and early printed texts, each `<pb/>` requires an `xml:id` value which (*qua* `xml:id` value) must be unique within the document. It also requires an `n` value, not necessarily unique, which is what will actually appear on users' screens. See the Element Set for how to assign these, and for the rules about spacing around `<pb/>`.

Variant procedure for modern printed texts

In modern printed texts (nineteenth century or later), no `xml:id` value is needed for `<pb/>`, just an `n` value.

6) In **Bible references**, if there's a full stop (or other punctuation mark) between chapter and verse, keep it, with no following space, e.g. '1 Iohn 5.7'. If there is no punctuation, leave a single space: 'Exodus 4 6'. But if more than one verse is referenced, leave spaces between the verse numbers, e.g. 'Dan. 6.9, 10, 11'. If uncertain about whether or not something is a Bible reference, or about what 'chapter' and 'verse' mean in Bible references, say so in a comment tag.

In non-Biblical references, you should normally leave a space after *any* punctuation in the original: e.g. 'Aug. de Civ. Dei 1. 3. c. 5' - *please* not '1.3.c.5' - this sort of mistake is very easy to overlook when proofing.

However, if you come across a case where the layout of the text makes it clear that there really is an intent on the author's or scribe's part to run the numbers and/or letters together, follow that intent. If in doubt, put in a comment tag.

7) Note that the only way to indicate **whitespace** in XML, whether horizontal or vertical, is by means of tags. The sole exception is the single space (one stroke of the spacebar). Longer spaces between words, line breaks, indentations and vertical gaps between paragraphs, etc, must all be indicated by tags. NEVER USE TABS TO INDICATE WHITESPACE.

The browser knows that line breaks are needed between paragraphs and lines of verse, and before and after headings. If a paragraph, line or heading is centered or irregularly indented, this must be indicated by the `rend` value of the relevant tag (`<p>`, `<l>` or `<head>`).

Any other **blank spaces in the text**, whether within a line of text or between two lines of text, are indicated by the `<space/>` tag. The `dim` value states whether the space is horizontal or vertical. The extent of the blank space is given as a numerical `extent` value and the type of unit being counted is specified in the `unit` value ("chars" (characters) for horizontal space, "lines" for vertical space). Note that `<space/>` may not occur between paragraphs or between `<head>` and `<p>`, `<lg>` or `<l>`. If there's a space between a heading and a paragraph or line of verse, or between two paragraphs, treat it as the last thing in the heading or in the first of the paragraphs:

```
<head rend="center">An historical account of two
notable corruptions of Scripture, in a Letter to a
Friend.<space dim="vertical" extent="2"
unit="lines"/></head>
<p>Sir</p>
<p>Since the discourses of some late writers ...
(New College 361(4))
```

8) If you come up against text arranged in **columns** or in **tabular form**, ignore it for now (but let us know about it) unless given individual instruction.

c) Characters

1) **Hyphens** at **line-breaks** and **carets** marking **insertions** should be recorded in the `<lb/>` or `<add>` elements. A hyphen at a line break is recorded by the `type` value: 'Baby<lb type="hyphenated"/>lon'. The presence or absence of a caret to mark an insertion is recorded by the mandatory `indicator` value of `<add>` (as "yes" if there is one or "no" if there isn't).

Carets that don't in fact point to anything can be ignored in the tagging, but feel free to mention them in a `<!--CODIC -->` tag. If a caret has been deleted but the insertion it points to hasn't, call it `indicator="yes"` and mention it in a `<!--CODIC -->` tag.

2) **Special characters** - that is, anything other than unaccented Roman letters, Arabic numerals (and Roman numerals provided they're also Roman letters), hyphens and these standard punctuation marks

. , ; : ? ! ' " () [] { } / \ ~

- all have to be represented by entities (see Entity Set). Do NOT use special fonts or keyboard toggle functions for these.

i) **Ampersands** (the character '&') should be represented by the entity `&`; (otherwise the browser will think they mark the beginning of an entity).

ii) **Letters with diacritics** (such as â, è, ö, ï, ç) and **digraphs** (or **ligatures**) (such as æ and Æ) should be represented by the entities listed in the Entity Set.

iii) Transcribe the letter **thorn** (the thing looking like a 'y' in 'y^e', 'y^t', 'y^m', etc.) as the entity `þ`. (In practice, this will seldom arise, since the vast majority of abbreviations involving thorns can be catered for by the entities `&the;`, `&that;` and `&them;` (see section a) 6 above). 'y^m' has to be hand-crafted, however, as it sometimes means 'than' and sometimes 'then'.) Although Newton's thorns in fact look exactly like his 'y's, the letter serves a completely different function and its use is potentially of considerable interest to language historians.

iv) Transcribe initial 'ff' - which functions like a modern capital 'F' - as the entity `&ff;`, for the same reason (this gives us the option of letting users choose whether to view it as 'ff', 'F' or something else again).

v) The various fancy characters Newton uses for **note indicators** or to indicate that a passage is to be **inserted from elsewhere in the manuscript** should be transcribed wherever they appear in the text, not edited out. They should be represented by the `<newtonSymbol/>` tag unless they are Roman letters, Arabic numerals, or anything included in the Entity Set (in which case use letters, numerals or entities). Provide a brief natural-language description of the symbol as the `value` value of `<newtonSymbol/>` (see the Element Set for guidance in assigning values).

vi) Characters that appear to function merely as **decorations** or **doodles** (e.g. a series of tildes filling up the space between the end of a line of prose and the right margin) can be mentioned in a `<!--CODIC -->` tag but should not be transcribed as such.

vii) **Greek and Hebrew characters** (including Greek characters with diacritics) are dealt with in the Entity Set, but transcribers unfamiliar with Greek and/or Hebrew are free to give up at this point and throw in a `<!--TODO NP -->` tag. **Greek digraphs and trigraphs** should be silently expanded but can be mentioned in comment tags if the transcriber feels so inclined.

3) Don't bother noting other distinctions between letter forms, such as long and short 's', Greek and Roman 'e', etc.

d) Formatting

1) **Superscript** and **underlining** are both indicated by <hi>, with the `rend` values "superscript" "underline" and "doubleUnderline". If **note or addSpan indicators** (whether represented as letters, entities or <newtonSymbol/>s) are placed above the line, tag them <hi `rend="superscript"`>, unless they've obviously been added as an afterthought, in which case tag them <add `place="supralinear"`>. If in doubt, put in a <!--TRANSC --> tag.

Sometimes when Newton wants to underline a lengthy passage, he saves time and ink by underlining only the first and last line of the passage and the first few letters of each line in between. In such cases, treat the entire passage as underlined, since this is clearly the intention, but mention the fact in a <!--CODIC --> tag, e.g. '<!--CODIC only first and last line of this passage and first few letters of the intermediate lines underlined - lc-->'.
<!--CODIC only first and last line of this passage and first few letters of the intermediate lines underlined - lc-->'

If underlining has itself been deleted, though the originally underlined words haven't been (i.e. Newton underlined a passage and then thought better of it), treat it as not underlined but note the fact in a <!--CODIC --> tag.

2) A **change of script** by the main author or scribe, or a change of font in printed text, e.g. to bold or italic, or to significantly larger script, should also be tagged <hi> (see the Element Set for permitted `rend` values). Don't bother recording slight fluctuations in the size of script, only mention them if there is a clear intent to make a particular word or phrase stand out by distinguishing it from the surrounding text.

3) A **change of hand**, i.e. where somebody else takes over the writing of the text, should be tagged <handShift/>, with the identity of the new hand indicated as the new value, using the codes listed in Name Codes. (NB: this is not necessary for insertions, deletions, notes or page numbers that appear in a different hand, as they are dealt with by <add>, <addSpan/>, , <note> and <fw>, each of which has its own hand attribute.) When (and if) the hand or style reverts to what it was before, say so in a second <handShift/> tag. For instance: 'passage of text in Isaac Newton's handwriting <handShift new="hn"/>which for some reason gets taken over by Humphrey Newton partway through <handShift new="in"/>and then reverts to Isaac's hand again.'

4) Tag **interlinear and marginal insertions** with <add>, provided this can nest within whatever element(s) it starts in. If it can't, use <addSpan/> (see section d) 5 below). Every <add> should have a `place` value as described in the Element Set, and an `indicator` value (of "yes" or "no") to state whether or not there is a caret or other indicator to mark the point of insertion. One <add> can nest within another if the layout of the text requires it. Do *not* mark up line breaks within <add>s, it will only cause confusion. If the insertion is in a **different hand**, indicate this using the hand value and the appropriate Name Code. Where one text string has been replaced by another by **overwriting**, indicate this using <del `type="over"`> and <add

place="over"> thus: 'B<del type="over">y<add place="over" indicator="no">i</add>th<del type="over">i<add place="over" indicator="no">y</add>nia' (Yahuda 15.7), meaning that Newton originally wrote 'Bythinia' but changed it to 'Bithynia' by overwriting the 'y' with an 'i' and the first 'i' of 'Bythinia' with a 'y'.

5) **Insertions from elsewhere in the text** (another page, or a different part of the same page), or **insertions that violate element boundaries**, should be transcribed where they belong in the text, introduced by `<addSpan/>` and terminated by an associated `<anchor/>`. Where such insertions appear on the main body of a page, line breaks *should* be tagged. The physical location of the passage is indicated by the `place` value of `<addSpan/>`. The procedure for assigning the `place` value and for hooking up an `<addSpan/>` to its associated `<anchor/>` is explained under `<addSpan/>` and `<anchor/>` in the Element Set. `<addSpan>` also has `startDescription` and `endDescription` values which provide, respectively, a short natural-language description of where the inserted text begins and where the main text is taken up again. These values are normally generated automatically and do not need to be entered by transcribers.

Even if the added section begins and/or ends on a different page from the main text, it should not be introduced or terminated by `<pb/>` (the function of that tag has been taken over by the `<addSpan/>` and `<anchor/>` tags), but if the inserted passage itself runs to more than one page, tag the page breaks within it `<pb/>` as normal. For instance:

```
<pb xml:id="p004r" n="4r"/>... Afric and Britain
being quieted a little before. <addSpan
spanTo="#addend003v-01" place="p002v p003v"
startDescription="f 2v" endDescription="f 4r"/>For
the history of the wars ... you may see in
Iornandes mention made of an incursion of the
Vandals out of Pannonia into Gallia: which Vandals,
as <pb xml:id="p003v" n="3v"/> the same Iornandes
relates, had been received into Pannonia by
Constantine ... the wars in Italy AD 536.<anchor
xml:id="addend003v-01"/> The first Trumpet begins
with the Visigothic wars ... (Yahuda 1.7).
```

This indicates that the text before the inserted section is on f. 4r, the insertion begins on f. 2v and continues on f. 3v from 'the same Iornandes' to 'the wars in Italy AD 536.', and then the text on f. 4r is taken up again with 'The first Trumpet begins'.

If this results in two or more `<pb/>`s having the same `xml:id` value (this is fairly unusual but it does happen), call the first one (for instance) `<pb xml:id="p034v-a" n="34v"/>`, the second `<pb xml:id="p034v-b" n="34v"/>`, and so on. It doesn't matter that the two `n` values are identical, but all `xml:id` values must be unique (*qua* `xml:id` values) within a document.

Newton often indicates the location of such inserted passages by beginning them with a **symbol** such as an obelus or a dot in a circle, and placing the same

symbol in the main text at the point he wants the insertion to appear at. These symbols should be recorded in the transcription, using entities or the `<newtonSymbol/>` tag:

```
<pb xml:id="p014r" n="14r"/> ... by degrees they
subdued it. <newtonSymbol
xmlns="http://www.newtonproject.sussex.ac.uk/ns/non
TEI" value="asterisk in a circle"/><addSpan
spanTo="#addend014v-01" place="p013v p014v"
startDescription="f 13v" endDescription="f 14r"/>
<newtonSymbol
xmlns="http://www.newtonproject.sussex.ac.uk/ns/non
TEI" value="asterisk in a circle"/> The calamity of
Afric in &the; first two or three years of this
invasion ... (Yahuda 1.7)
```

When Newton doesn't include such indicators, it can be very difficult to decide where exactly a supplementary passage does belong, or, indeed, whether Newton himself was entirely sure where he wanted it to go. When in doubt, add a `<!--TRANSC -->` tag, e.g. `<!--TRANSC not sure if this belongs here - jy-->`. Do make sure that everything in your document gets transcribed (apart from passages you can't do and have clearly assigned to someone else in a `<!--TODO -->` tag), even if you have no idea where some of it belongs - again this can be pointed out in a `<!-- -->` tag, e.g. `<!--TODO NP somebody figure out where this addSpan belongs, I can't make any sense of it - jy -->`. One option in such cases, to avoid interrupting the flow of your main text, is to put the 'orphaned' `<addSpan/>` passage(s) at the very end of your transcription, with the physical location of each passage noted in the `place` value and an `endDescription` value of "unknown": it will then be down to the editors to decide where to put it in the final version.

The distinction between `<addSpan/>` (supplementary text) and `<note>` (annotation) can be very difficult to draw, especially when Newton is in one of his less coherent moods. If in doubt, say so in a `<!--TRANSC -->` tag.

6) Where two or more **alternative readings** are placed one above another, use the `<app>` and `<rdg>` tags as described under `<rdg>` in the Element Set.

7) **Deletions** should be indicated by ``. Distinguish four levels of deletion, using these `type` values: "blockStrikethrough" for whole sections struck through en bloc (usually by a large cross or a diagonal line), "strikethrough" for a text string crossed out with a continuous horizontal line, "cancelled" for any heavier deletion than that, and "over" for words or characters that have been overwritten by other words or characters (see section d) 4 above).

It is perfectly alright for one deletion to nest inside another. This makes it possible to indicate multiple revisions. For instance, in Yahuda 1.5 f. 60v, there's a passage where Newton had obviously first written 'the Franks ... were up in arms before', then changed that to 'the Franks ... were in a posture of war before the rest, & that with so great force as to Conquer the Conquerors', and then crossed almost the whole lot out and rewritten it as 'the Franks ... were up in arms before the rest &

animated with victory over the victors of the Romans'. Later still he struck out the entire passage en bloc. This yields:

```
<del type="blockStrikethrough">the Franks ... were
<del type="strikethrough"><del
type="cancelled">up</del> in a<del
type="cancelled">rms before</del> posture of war
before the rest, & that with so great force as
to Conquer the Conquerors</del> up in arms before
the rest & animated with victory over the
victors of the Romans.</del>
```

When a deleted text string has been replaced by an **inserted** one, try in general to place the `<add>` *after* the ``, thus reflecting the order in which the text was actually written, even if the caret mark (if there is one) appears, physically, before the ``. Don't worry if this means skipping back a line to start the addition. The following is perfectly acceptable:

```
<del type="strikethrough">But when Alaric had
<lb/>thus sufficiently derided the lost condition
of the Empire he degrades <lb/>Attalus &</del>
<add place="supralinear" indicator="yes">But
Attalus behaving him self foolishly Alaric degrades
him again &</add> restores Honorius (Yahuda 1.7)
```

even though 'But Attalus behaving him self foolishly' is in fact written above the deleted 'But when Alaric had', i.e. two lines higher than where it appears in the transcript.

However, if a chunk of text has been deleted and replaced by an `<addSpan/>` passage, transcribe the `<addSpan/>` passage first, especially if it has an indicator located, physically, before the deleted section. And if an insertion and a deletion occur at more or less the same point in the text but appear to be quite independent of one another (i.e. the inserted text string does not replace the deleted one), transcribe the insertion where its caret occurs if it has one, and otherwise transcribe them in whatever seems the most logical order: for instance, 'the great `<add place="supralinear" indicator="no">red</add>` `<del type="strikethrough">Beast` Dragon'.

The spacing before and after `<add>` and `` tags should generally be exactly what it would be if the tags were not there: thus '`<del type="word strikethrough">his` `<add place="supralinear indicator="no">her</add>` child', but '`supers<add place="supralinear" indicator="yes">ti</add>`tion' and 'the `<del type="cancelled">Ostro<del type="over">g<add place="over" indicator="no">G</add>`oths were up in arms'.

`<add place="over">`, however, should always follow `<del type="over">` *immediately*, with no space between them, even when a whole word has been overwritten.

Where text is deleted by **another hand**, indicate this using the hand value of `` and the appropriate Name Code. Obviously, identifying the hand of a deletion is often difficult or impossible, especially if working from a black-and-white microfilm printout that doesn't reveal changes of ink, but in a number of texts where (for instance) Newton has corrected the work of an amanuensis, it is clear that some at least of the deletions were done by him and not by the original scribe.

If deleted passages violate other element boundaries (this is a fairly rare occurrence), for instance if a paragraph and a half or two whole paragraphs have been struck through with a single cross, or if a deletion begins in the middle of an insertion and then carries on into the main text, treat it as two or more separate ``s, but mention the fact in a `<!--TRANSC -->` tag.

It's not unknown for one `<del type="blockStrikethrough">` to nest inside another, if, for instance, Newton has a big X taking out three lines of text and a bigger X taking out the entire page within which those three lines occur.

8) **Wholly illegible deletions** should be tagged `<gap reason="illgblDel"/>` (with, obviously, the appropriate `type` value for ``), and their length indicated by the `extent` and `unit` values of `<gap/>`. `<gap reason="illgblDel">` should, by definition, only ever occur within ``. However, if the illegible bit nests in a longer deleted but otherwise legible passage, it does not need a special `` tag of its own. For instance:

```
<del type="strikethrough">Probably Pelasgus from
whom they had their name was one of the sons of
<gap reason="illgblDel" extent="1" unit="word"/>
that Elisha who first peopled Peleponnesus.</del>
(Keynes 146)
```

9) **Annotations** that appear in manuscripts require an `<anchor/>`, which appears at the point the annotation refers to. The annotation itself is tagged `<note>`, with a `place` value indicating where, physically, it appears. It should be transcribed immediately after its `<anchor/>`, leaving no space in between. The `<anchor/>` has an `xml:id` value which must be unique within the document and which is pointed to by the `target` value (or one of the `target` values) of the `<note>` it refers to. The system for assigning these values is explained in the Element Set under `<anchor/>`.

If the text includes its own indicator of the point to which an annotation refers, such as a symbol or a superscript letter, place the `<anchor/>` immediately after the indicator. As with `<addSpan/>` passages (see section d) 5), these indicators should be included in the transcription, not edited out. If there is no such indicator, put the `<anchor/>` at what seems the most appropriate point, leaving no space between it and the point it refers to, e.g. 'as Augustine<anchor xml:id="n006r-03"/><note place="marginRight" target="#n006r-03">De Civ. Dei l. 8 c. 4</note> saith'. This is usually pretty obvious from the context, but if in doubt, mention the problem in a `<!--TRANSC -->` tag.

It is sometimes necessary to link more than one `<anchor/>` to the same `<note>`, as when Newton adds, say, a note 'b' to a passage, and then redrafts the passage with another note 'b' indicator at the relevant juncture, but doesn't bother to

copy the note out all over again. In such a case, each note indicator (the letter 'b') is immediately followed by an `<anchor/>` with a unique `xml:id` value, and the `<note>` itself, which should be transcribed immediately after the first `<anchor/>`, takes two `target` values, separated by a single space. Thus, the first note indicator might be followed by `<anchor xml:id="n053r-01"/>` and the second by `<anchor xml:id="n053r-02"/>`, while the note itself is tagged `<note place="marginRight" target="#n053r-01 #n053r-02">`.

Notes that have been **deleted** should have a `` tag nested inside them, irrespective of whether they themselves nest in a deleted passage or not. The same principle applies to underlining and any other formatting within notes, i.e. a `<hi>` tag is needed within the `<note>` irrespective of any formatting on the text in which it nests.

If the annotation is in a **different hand** (e.g. where John or Catherine Conduitt has annotated a Newtonian manuscript), this should be recorded as the `hand` value of the `<note>` (see Name Codes).

If a note runs to more than one page, it may contain `<pb/>` and `<fw>` as appropriate. If this results in two or more `<pb/>`s having the same `xml:id` value, modify the `xml:id` values by adding -a, -b etc., as with `<addSpan/>` (see section d) 5 above).

Variant procedure for printed texts

Annotations in printed texts are much simpler to deal with. `<anchor/>` is not needed, and instead of a `target` value the `<note>` has an `n` value which is simply the letter, figure or symbol that functions as the note indicator in the original text. If it is a symbol, encode it as an entity, e.g. `<note n="&obelus;">` (see Entity Set). The note indicator itself should not be transcribed, merely recorded as an `n` value. If there is no note indicator in the original text, no `n` value is needed: simply transcribe the note at the point to which it refers. If this is not clear, mention the fact in a `<!-- TRANSC -->` tag. No `place` value is needed for `<note>`s in printed text.

Element Set

In the following, attributes that are *mandatory* (i.e. the transcription will not validate unless they are included) are preceded by **. Attributes that are not mandatory according to the schema but are nonetheless necessary for the transcribed text to display properly are preceded by *. Other attributes can (and should) be omitted where there is no call for them.

Some elements and attributes will not normally be entered by transcribers but are listed here so that you know what they are if you encounter them when proofing someone else's work.

`<!-- -->` [Comment.]

No attributes. This is for informal in-house comment - messages from the transcriber to him/herself and/or other project members. Such comments fall within five categories and should begin with the appropriate code (so that the editors can more easily identify the ones likely to require their personal attention).

`<!--TODO -->`: indicating a problem you can't solve but are fairly sure someone else can. `TODO JY` means you think John Young can solve it, `TODO MJH` means you think Mike Hawkins can, and `TODO NP` means someone on the project needs to look at it but you're not sure who. You can also assign `TODOS` to yourself, for things you can't face at the time but want to come back to later.

`<!--TRANSC -->` (transcription): notes on any uncertainties or reservations you have about your own transcription and/or encoding.

`<!--CODIC -->` (codicological): to point out features of the manuscript that these guidelines don't yet provide any means of encoding, e.g. '`<!--CODIC this page written upside down - jy -->`', '`<!--CODIC the bottom of the page has been cut off, it isn't clear if there was any text on it - jy -->`'.

`<!--APP -->` (apparatus): notes that may be of use to the editors when they come to add apparatus, e.g. recognising the source of an unattributed quote, spotting an error in a Biblical reference, or anything else where you think your expertise might contribute to the apparatus.

`<!--OTHER -->`: anything else.

Such comments are not treated as part of the text and (obviously) will not appear in the online or any printed version. It's a good idea (though not compulsory) to initial such notes unless they are purely for your own use and to be deleted after you've used them. This means that if the document is passed on to someone else, they know who to contact if they have any answers or suggestions. If when proofing a document you resolve a problem flagged by someone else in a `<!-- -->` tag, delete the comment.

This tag is particularly useful for the editors when preparing a document for release, as it provides an easy way of identifying points where

transcribers have encountered difficulties that may need editorial attention. **It cannot be overstressed that admitting to such difficulties is no cause for embarrassment. It will be looked on as evidence not of incompetence but of commendable honesty.** Besides, it may very well be that the problem lies with the guidelines themselves rather than your failure to understand them, and your comments will provide useful input for the next upgrade.

<abbr> [This element is used in tandem with **<expan>** and both are contained within a **<choice>** element. **<abbr>** contains any conventional abbreviation other than a brevigraph (for which see **<orig>**). **<expan>** contains the expanded form of the text.]

Notes

The commonest conventional abbreviations - 'y^{er}', 'yth', 'w^{chr}', 'wth' and many more - can be encoded as entities (see Entity Set). Don't bother tagging abbreviations that are still conventional, e.g. 'Dr', 'Mr', 'lib.' (for 'liber'). Don't tag abbreviations of proper nouns, e.g. 'Matt.' for 'Matthew'. And don't tag idiosyncratic abbreviations like 'i n o t F t S a t h g', but do feel free to throw in a comment tag such as '<!--APP I guess he means 'in the name of the Father, the Son and the holy ghost'-->'.

Examples

```
'His <choice><abbr>Ma<hi
  rend="superscript">tie</hi></abbr><expan>Majest
  ie</expan></choice>'
'<choice><abbr>S<hi
  rend="superscript">tus</hi></abbr><expan
  cert="high">Spiritus</expan></choice>'
```

<add> [Addition or insertion of a text string, whether by the author or in another hand.]

Attributes

- ** *indicator*: if there is a caret mark or other indicator to specify the point of insertion, this takes the value "yes", and if not, "no".
- * *place*: where, physically, the addition occurs. Permitted values: "supralinear" (above the line), "infralinear" (below the line), "inline" (a text string squeezed in between but not above or below two others), "interlinear" (inserted text running to more than one line so that it can't be adequately described as either above or below the point of its insertion), "lineEnd", "lineBeginning", "marginRight", "marginLeft", "over" (for a text string that replaces another by overwriting it). "lineEnd" and "lineBeginning" are for text added in the margin at the end or beginning of a given line of text; "marginRight" and "marginLeft" are for more substantial marginal insertions or for text that is written in the margin but is clearly intended to appear part way through a given line. These can be combined in the appropriate order, separated by single spaces, if an addition sprawls over different places, e.g. '<add place="inline interlinear marginRight" indicator="no">' means the addition starts inline, continues above the line and then proceeds into the right margin. If these values

aren't adequate, e.g. if the added material is wholly or partly on another page, the addition should be dealt with using `<addSpan/>`.

hand: code for the person doing the addition (see Name Codes) if it's in a different hand from that of the surrounding text. Not needed for additions in the same hand as the surrounding text (which are normally the overwhelming majority). If uncertain about the identity of a hand, say so in a `<!-- TRANSC -->` tag.

cert: degree of certainty about whether the text string is in fact an addition on a scale of "high", "medium" or "low". Normally there is no doubt so no *cert* value is needed, but where text runs into the right margin (for instance) it can be unclear whether it was added as an afterthought or the writer simply overshot the margin.

Notes

If an `<add>` is wholly deleted, nest the `` immediately inside it: 'whose worship the Prophets upbraid with folly by representing that the Idols `<add place="supralinear" indicator="yes"><del type="strikethrough">`of the`</add>` can neither hear nor see nor walk' (Keynes 7). But `<add>` may also nest inside `` if the layout of the text demands it: 'his 130th Ep. `<del type="strikethrough">`written to Symplicius (as Gothofredus thinks `<add place="supralinear" indicator="no">`as Goth. thinks)`</add>`, when Master of the Hors``' (Yahuda 1.6).

`<add>` can also nest within `<add>` if an inserted passage itself contains a further insertion. In such cases, the *place* value refers to where the nested insertion appears relative to the inserted text it nests in.

An `<add>` that replaces a deleted word or passage should be placed after the deletion, with one space between the closing `` and the opening `<add>` tag, unless it has the *place* value "over", in which case it should follow without a break. If it replaces a deleted letter or letters that are only part of a word, no space should be left, e.g. 'a`<del type="cancelled">d<add place="inline" indicator="no">l</add>`lowed', 'return`<del type="strikethrough">ing<add place="supralinear" indicator="no">ed</add>`'. (See *The Text* section d) 7 for further guidance on placement of `<add>` and spacing around it.)

Must nest within any element it starts in and may not contain `<p>`. If an entire paragraph has been squashed in between two pre-existing ones, nest `<add>` directly inside `<p>`, with the *place* value "interlinear". Additions that violate element boundaries should be dealt with using `<addSpan/>`.

`<addSpan/>` [Beginning of a section of inserted text from elsewhere on the page or in the Ms., and/or which violates element boundaries and therefore cannot be rendered as `<add>`. The inserted text begins at the `<addSpan/>` and continues to the corresponding `<anchor/>`.]

Attributes

** *spanTo*: identical value to `xml:id` of relevant `<anchor/>` (q.v.) prefixed by a hash character (#).

* *place*: where, physically, the addition appears. Values as for `<add>`, or, if the added text is wholly or partly on a different page from its point of insertion, the `xml:id` value of the `<pb/>` of the page in question. If it runs to more than one page, list them all, separated by single spaces: `<addSpan spanTo="#addend023v-01" place="p021v p022v p023v" startDescription="f 21v" endDescription="f 22r"/>`. If it is on a discrete section of the page, this can be indicated by e.g. `place="p064r-marginRight"`, `place="p064r-lower"`, `place="p064r-higher"`. If uncertain how to describe the location (and it can often be far from straightforward), explain the problem in a `<!--TRANSC -->` tag. See *The Text* section d) 5 for more detail on how to deal with such passages.

hand: code for the person who's done the addition if it isn't in the same hand as the text into which it's to be inserted (see Name Codes). Not needed if it is in the same hand (as it usually is). If uncertain about the identity of a hand, say so in a `<!-- TRANSC -->` tag.

cert: degree of certainty about whether the passage really is an insertion (as opposed to a `<note>` or a normal section of text, on a scale of "high", "medium" or "low". Not needed if there is no doubt, as there usually isn't.

startDescription (not normally entered by the transcriber): a succinct natural-language statement of the point at which, physically, the addition begins, e.g. "f 21v", "p 5", "lower down f 17r", "right margin of p 7" ('f' if the document is being numbered by folio, 'p' if it's being numbered by page). Like the `n` value of `<pb/>`, this is what users will see on their screens at the relevant point.

endDescription (not normally entered by the transcriber): a succinct natural-language statement of the point at which the main text resumes (i.e. where you were immediately *before* starting the `<addSpan/>`): again, this is what users will see on their screens.

The values of *startDescription* and *endDescription* will be generated automatically by various executable scripts on our server. It is consequently not necessary for you to ever enter information into these fields except in the most exceptional of circumstance (approved in advance by the chief editor).

Notes

The `<addSpan/>` procedure is probably the most complicated thing in these guidelines, and takes a lot of getting used to. Don't be disheartened if you get it wrong the first few times.

It's not unusual for one `<addSpan/>` passage to start in the middle of another, which is when it gets really confusing, but the instructions above still apply. If in such a case both `<addSpan/>` passages end at the same point, they still need to have an `<anchor/>` each, the `<anchor/>` for the first `<addSpan/>` occurring immediately *after* that for the second `<addSpan/>`, so that the second passage 'nests' in the first.

<anchor/> [Point of reference for a manuscript <note>, or marker for the end of an <addSpan/> passage.]

Attributes

* *xml:id*, which is generated as follows:

- a) If the <anchor/> terminates an <addSpan/> passage:
take the *xml:id* value of the <pb/> of the page on which the anchor occurs and replace the initial 'p' with 'addend'. Then add a hyphen and a two-digit number. The first occurrence of an <addSpan/> anchor on a given page takes the suffix -01 (even if it is the only one on the page), the second takes the suffix -02, the tenth (in the unlikely event of that many ever occurring) takes -10, and so on. So the second instance of an inserted passage ending on f. 64v is marked by
`<anchor xml:id="addend064v-02"/>`.
- b) If the <anchor/> is serving as a note indicator:
take the *xml:id* value of the <pb/> of the page on which the note is referenced, replace the initial 'p' with 'n', and add a hyphen and a two-digit number as with <addSpan/> anchors. Thus, the fourth note referenced on f. 79r requires an <anchor/> with the *xml:id* value "n079r-04". This is the number of the page on which the <anchor/> occurs, even if the note itself is on a different page.

Notes

If the text includes its own note indicator (such as an asterisk or superscript letter), it should be transcribed using letters, numerals or entities where possible, or otherwise as a <newtonSymbol/> (q.v.). The <anchor/> should appear immediately after it, without a space between. Otherwise, place the <anchor/> immediately before or after the word or passage you think the note refers to. Keep it as close as the sense permits to where the note actually occurs. For instance, if Newton quotes a paragraph from Plutarch with a note reading 'Plutarch l. 3' in the margin parallel to the beginning of the passage, and does not provide his own note indicator, put the <anchor/>, followed by the <note>, immediately *before* the passage, not at the end as you would with a printed footnote.

When deciding which two-digit suffix to give an <anchor/>, be guided by where it appears *in the transcribed text*. Notes and <addSpan/> passages quite often appear, physically, in the manuscript in a different order from that in which they are referenced in the transcribed text. For instance, Newton may start adding marginal notes about halfway down the page, but run out of space by the time he gets to the sixth note and so put it at the top instead. The <anchor/> should still be given the suffix '-06' and the note be transcribed immediately after it. Or the first reference to a passage to be inserted from f. 28v may be to what physically appears as the third paragraph on f. 28v, with the first two being marked for insertion later on in the main text - but the <anchor/> still takes the value *xml:id*="addend028v-01". You can, of course, comment on such anomalies in a <!--CODIC --> tag.

If you find that your `<anchor/>` numbers have got out of kilter (for instance because you've changed your mind about where in the text an `<addSpan/>` passage or a `<note>` belongs, or you realise you've missed one out), don't lose sleep. It doesn't really matter what order the suffixes are in, so long as the `xml:id` values are all unique within the document and link the right `<anchor/>`s to the right `<addSpan/>`s or `<note>`s.

Examples

```
'About seven years after that captivity when
Sennacherib warred in Syria, he sent this
message to the King of Iudah. <anchor
xml:id="n004r-02"/><note place="marginRight"
target="#n004r-02">2 King. 19.11.</note>Behold
thou hast heard what the kings of Assyria have
done ...' (Keynes 146)
```

```
'&ff;or Pope <hi rend="superscript">o</hi><anchor
xml:id="n111v-01"/><note place="marginLeft"
target="#n111v-01">o Epist. apud Athanas. Apol.
2</note> Iulius tells us' (Yahuda 15.6)
```

See *The Text* section d) 5 for examples of `<anchor/>` terminating `<addSpan/>` passages.

<app> See **<rdg>**

Attributes

** *type*: either "authorial" or "variantTexts". "authorial" means the author or scribe has clearly indicated a choice of two or more alternative readings *in a given text*. "variantTexts" means that the editor is noting a variant reading *in another draft or edition of the same text*. At present, transcribers will not normally be required to deal with `type="variantTexts"`, and will certainly not be required to without receiving individual instruction.

<choice> See **<abbr>**, **<orig>**, **<sic>** and **<unclear>**. The function of the **<choice>** element is to link two or more alternative renditions of a single text string.

Note

`<choice>` may nest within `<choice>`, so if you have (for instance) two alternative readings of an abbreviated word, each of which would lead to a different expansion, they can be tagged as in the final example below.

Examples

```
'<choice><abbr>Doct.</abbr>
<expan>Doctor</expan></choice>'
'<choice><orig>&aover;</orig><reg>am</reg></choice>'
'<choice><unclear reason="copy"
cert="medium">dump</unclear><unclear
cert="medium">bump</unclear></choice>' (i.e. the
transcriber is unsure whether the text reads 'dump' or 'bump' and does
not regard either as likelier than the other)
```

'<choice><unclear reason="del"
 cert="high"><choice><abbr>w<hi
 rend="superscript">ch</hi></abbr><expan>which</
 expan></choice></unclear><unclear reason="del"
 cert="low"><choice><abbr>w<hi
 rend="superscript">t</hi></abbr><expan>what</ex
 pan></choice></unclear></choice>' (i.e. the transcriber is
 unsure whether text a) reads 'w^{ch}', in which case it should be expanded
 as 'which', or b) reads 'w^t', in which case it should be expanded as
 'what', and he/she strongly favours the former option while not entirely
 ruling out the latter)

<corr> See <sic>

 [contains text deleted by the author or another scribe.]

Attributes

** *type*: describes the level of deletion. Permitted values: "blockStrikethrough"
 (whole section struck through en bloc), "strikethrough" (continuous
 horizontal line through a text string), "cancelled" (heavier deletion);
 "over" (one text string overwritten by another).

status: where material has been deleted that should not have been, takes the
 value "erroneous".

hand: if the deletion has obviously been made by another hand, indicate this
 with the appropriate Name Code. Not needed if the deletion is in the
 same hand as the original text string (as it usually is).

Notes

The borderline between "strikethrough" and "cancelled" is somewhat subjective
 but not really all that important.

Must nest within any element(s) it starts in, and may not contain <p>. If a
 deletion does violate element boundaries, simply treat it as two (or
 more) separate s, but add a <!--TRANSC --> tag for future
 reference (see the third example below and *The Text* section d) 7).

 can nest within if a text string in a deleted passage had itself
 already been deleted before the rest of the passage was.

If the reading of the deletion, or part of it, is uncertain, nest an <unclear
 reason="del"> tag in the . If the deletion, or part of it, is
 completely illegible, nest <gap reason="illgblDel"/> within
 the .

Examples

'Valentinian II <del type="cancelled">by the <gap
 unit="words" extent="2"
 reason="illgblDel"/> sided with the
 Arrians in Italy' (Yahuda 1.5)

'Let us hear Claudian who brings in Rome thus <del
 type="cancelled">speaking to <unclear
 reason="del" cert="low">I</unclear>
 supplicating Iupiter' (Yahuda 1.6)

'<p><del type="blockStrikethrough"> ... as was
 analogous to the pain of death.</p>

`<p><del type="blockStrikethrough"><!--TRANSC
actually, the strikethrough crosses the
paragraph break - jy-->I know not whether it
may be worth </p>' (Yahuda 1.7)`

`'These last words no doubt relate to the old
unprofitable fables of those <del
type="over">that<add place="over"
indicator="no">who</add> lived according to the
law' (Yahuda 15.6)`

`'that answer &which; S<hi rend="superscript">t</hi>
<del type="cancelled">I<add
place="supralinear"
indicator="yes">Hi</add>erome gives' (Keynes 2)`

`'Saturn, Iupiter, Venus, <del
type="strikethrough">Mercury<del
type="strikethrough" status="erroneous">
and Mars'`

<div> [Self-contained section such as a chapter or sub-chapter.]

No attributes. Note that `<div>`s can nest within other `<div>`s (e.g. chapters within a document or sub-chapters within chapters). However, `<div>`s must 'tessellate': that is, one `<div>` must end before another can start. This is OK:

```
'<div><div> ... </div><div> ... </div></div>'
```

but this is not:

```
'<div> ... <div>...</div> ... </div>'
```

Because nobody told early modern writers about 'tessellation', self-contained sections do in fact quite often 'float' within their parent `<div>` in this fashion, but they can't be regarded as `<div>`s. Simply transcribe the 'floating' section as though it were an undifferentiated part of the text, but if it has a heading, tag the heading `<floatingHead>` (q.v.).

If the document has no sub-sections, the entire text goes in a single `<div>` nesting directly in `<body>`.

<expan> See **<abbr>**

<figure> ["groups elements representing or containing graphic information such as an illustration or figure." From TEI]

Transcribers will not be expected to enter this or `<figDesc>` unless given special instruction. Simply note the presence of a figure in the text in a `<!-- TODO NP -->` tag, e.g. `'<!-- TODO NP diagram in the right margin -->'`.

<figDesc> [Brief prose description of a graphic figure.]

<figDesc> nests within the <figure> it describes (it does not particularly matter where).

<floatingHead> [Something that behaves in all respects like a <head> except that it does not in fact introduce a new <div> or <lg>.]

Attributes

rend: as for <head>

Notes

This can be used for the headings of sub-divisions that 'float' in the middle of their parent divisions and therefore can't be tagged <div> (see above under <div>). Another example is Newton thinking he's come to the end of a sub-section, writing a heading for the next sub-section, then thinking of something he wants to add to the first sub-section, deleting the heading and resuming the text where he'd left off.

<foreign> [Text in a different language from the text immediately surrounding it.]

Attributes

* *xml:lang*: language. Permitted values: "ara", "eng", "fre", "gre", "heb", "lat" (Arabic, English, French, Greek, Hebrew, Latin). If you come across anything else, consult an editor.

Notes

Must nest within whatever element(s) it starts in and may not contain <p> or <l>, which of course foreign language passages frequently do: in such cases, use repeated <foreign> tags nesting in <p> or <l>. If, as sometimes happens, the document simply changes language halfway through, or goes into a different language for pages on end, don't bother nesting <foreign> in every single <p> or <l>: just mention the fact in an <!--APP --> note. But if in doubt, apply the tag: it can't do any harm.

Note that English may be regarded as a foreign language if the main language of the document is something else (usually Latin).

<formula> [Container element for all mathML encoded passages.]

On coding **presentational** mathematical content, see:

<http://www.w3.org/Math/testsuite/mml2-testsuite/index.html>.

Another useful resource is: <http://www.mathmlcentral.com/>

The bulk of the coding in presentational mathML will *likely* be done with a very small subset of the available tags, chief of which will be:

<mfrac>

<mi>

<mn>

<mo>

<mtext>

<mrow>

<root> / <msqrt>

<msup>

Transcribers dealing with mathematical texts will be given special instruction on how to apply these.

<fw> [Forme work, i.e. scribal or printed features of the document that don't constitute part of the text proper, such as page numbers and catchwords.]

Attributes

* *type*. Permitted values: "pag" (page number), "catch" (catchword), "header" (running title at head of page), "sig" (sigil), "shelfmark".

place. Not needed for headers. Combinations of "top", "bottom", "Left", "Right" and "Center": page nos. are usually "topRight" but sometimes "centerRight" or "topCenter". We don't need to be too precise about this. Catchwords are nearly always "bottomRight", but sometimes the last word on a page *functions* as a catchword (i.e. is duplicated at the beginning of the following page), although it isn't in the normal catchword position: in such a case call it `place="inline"`.

hand: code for the person who's added the forme work (this is most likely to apply to page numbers and shelfmarks) if it isn't in the same hand as the surrounding text (see Name Codes).

Notes:

Not required in transcriptions of modern printed documents.

If a paragraph ends at the bottom of a page, and the catchword is duplicated by the first word of the next paragraph, the catchword should go between the closing `</p>` tag and the opening `<p>` tag, as should the `<pb/>` and the page number (if there is one) on the new page.

Some Mss. have two or more differing page numbers on some or all of their pages: each page number should have its own `<fw>` tag, but if this happens mention it in a `<!--CODIC -->` tag. No spaces should be left between the consecutive `<fw>` elements.

Hyphens on partial catchwords, or on partial words preceding catchwords, should be recorded as `<lb type="hyphenated"/>`, though `<lb/>` is not otherwise needed at these points.

No space should be left between `</fw>` and `<pb/>` or between `<pb/>` and `<fw>`.

<gap/> [Material not transcribed (because illegible or lost).]

Attributes

** *reason*: why it's missing from the transcription. Permitted values: "illgblDel" (illegible deletion), "copy" (poor quality of the copy you're transcribing from), "damage" (manuscript damage, e.g. it's torn or has a hole in it -- indicate the precise nature of the damage in the `agent` attribute), "blot" (blotted), "blotDel" (text is obliterated by what could either be an accidental blot or a deliberate deletion), "smudge", "over" (text is impossible to read because it is itself written over other text - if it's impossible to read because other text is written over it, it counts as "illgblDel"), "faded", "foxed" (manuscript has gone brown or crozly round the edges through age or damp), or just "hand" if it's simply a case of bad handwriting, though normally a guess at least should be possible, tagged `<unclear reason="hand" cert="low">`. If none of these values seems to fit exactly, use whichever you think comes closest and specify the problem in a `<!--TRANSC -->` comment.

agent: if the text is illegible because of damage, you can use this attribute to provide further information on the 'agent' or 'cause' of the damage (e.g. "rubbing", "fire", "mildew").

extent and *unit*. These two variables are used together to indicate the quantity of the illegible text. *extent* contains the numeric value and *unit* contains the unit of measurement ("words" or "chars", i.e. characters). Note that the *unit* value is always plural, even if only one word or character is missing or illegible. Alternatively, *extent* can have the value "unclear" (e.g. if a page has been torn across the middle and there is no way of even guessing how much text has been lost), in which case no *unit* value is needed.

Notes

Does *not* mean 'blank space in the text', which is dealt with by `<space/>`.

Leave a single space before and after `<gap/>`, unless it represents only part of a word, or begins and/or ends part way through a word, in which case there should be no space between `<gap/>` and the legible part(s) of the word(s).

`<handShift/>` [Change of hand.]

Attributes

* *new*: code of person whose hand begins at this point (see Name Codes).

resp: indicates the encoder/editor responsible for determining that a change of hand has occurred and identifying the new hand. The value of this attribute will be the `xml:id` value of the encoder/editor (as defined in the header) prefixed by a hash character (#).

Notes

Obviously, when (and if) the hand reverts to what it was before, this has to be pointed out by another `<handShift/>` tag. See *The Text* section d) 3.

Not needed for `<add>`s, `<addSpan/>`s, `<note>`s or `<fw>` that have been added by a different hand, or for deletions made by a different hand, as this is expressed by the *hand* value of the relevant tag.

`<head>` [Heading of a `<div>` or `<lg>`.]

Attributes

rend: "center" if the heading is (more or less) centered. If it isn't, no *rend* value is needed.

Notes

Don't worry if a heading isn't *exactly* centered. The layout of the text is the least of our worries, and people who are interested in it will want to look at the original or a facsimile anyway. We don't need to give more than a general idea.

Users of British English please note the American-cum-Newtonian spelling of 'center'.

Example

```
'<div><head rend="center">Paradoxical Questions
concerning the morals & actions of
Athanasius & his followers.</head>
<div><head rend="center">Quest. 1. <lb
```

```
type="intentional"/>Whether the ignominious
death of Arius in a bog-house was not a story
feigned & put about by Athanasius above
twenty years after his death.</head> ... </div>
... </div>' (Keynes 10)
```

[Note here the nested <div>: 'Paradoxical Questions' is the title of the entire document, so the first <div> will not be closed till the end of the document; 'Quest 1' is the first section, so the second <div> will be closed at the end of that section and a new <div> opened for Quest. 2.]

<hi> [Text rendered graphically distinct in some way (other than by deletion).]

Attributes

**** *rend*:** what makes it distinct. Permitted values: "superscript", "subscript", "underline", "doubleUnderline", "overline" (see <orig> for the use of this), "bold", "italic", "large", "larger", "largest" (if text becomes significantly larger), or "small", "smaller", "smallest" (if the text becomes significantly smaller).

Notes

Must nest within whatever element(s) it starts in and may not contain <p>.

This is seldom a problem, but if you come across something like two continuously underlined paragraphs, nest a separate <hi> tag in each of them (cf.).

Tag changes of text size sparingly, for words or phrases that are clearly being deliberately highlighted, not for places where the writing gets a bit larger or smaller just because Newton was tired or excited or getting cramp. In practice, "larger", "largest", "smaller" and "smallest" seldom if ever occur in manuscripts (their main use is in printed title pages).

<l> [Line of verse.]

Attributes

***rend*.** Permitted values: "center", or "indent*", * being replaced by the approximate number of spaces the line is indented, in multiples of 5 up to a maximum of 40. Not needed if it isn't indented at all with respect to the other lines around it.

Note

Verse passages should normally consist of <l>s nesting within <lg> (line group), even if they are only one line long.

Neither <l> nor <lg> may nest within <p>, for some reason. If a verse passage occurs within a paragraph, put a closing </p> tag before the verse passage and then an opening <p *rend*="indent0"> tag after it.

<lb/> [Line break in prose sections.]

Attributes

***type*.** Permitted values: "intentional", for where we believe that the break has been intentionally introduced because it is or could be construed as structurally or semantically significant (usually in <head>). The other permitted value is "hyphenated", used when a hyphenated line break

occurs in the middle of a word. In all other cases (including line breaks that occur mid-word where there is no hyphen in the original), no `type` value is needed.

Notes

`<lb/>` is not needed at the beginning or end of a page, heading, paragraph, line of verse or verse passage, or after an unhyphenated catchword.

Unless nesting in `<fw>`, `<lb/>` should be placed right at the beginning of the following word (or part of a word) but have before it whatever spacing would be there if `<lb/>` weren't (so if it's in the middle of a word, no space, if it's at the beginning of a word, one space: see *The Text* section b) 3).

Use only in text on the main body of the page, including `<addSpan/>` passages and `<note>`s that occur on the main body of a page, but *not* in marginal or interlinear additions or notes.

Examples

```
'&ff;or by what we produced above out of
  <lb/>Claudian, it is manifest that all Gallia
  & Spain conti<lb type="hyphenated"/>nued
  quiet till that great irruption of barbarians
  which be<lb type="hyphenated"/>gan the wars of
  the second Trumpet' (Yahuda 1.5)
```

```
'<head rend="center">Quest. 2. <lb
  type="intentional"/>Whether the Meletians
  deserved <lb/>that ill character &which;
  Athanasius <lb/>gave them</head>' (Keynes 10)
```

[In the second example, there is a clear intention to force a line break between 'Quest. 2.' and the rest of the title, but the other line breaks occur only because that's where Newton happened to reach the end of the line, and so should not be tagged `type="intentional"`.]

<lg> [Line group, i.e. a verse passage or a stanza within a verse passage.]

Attributes

None

Notes

`<lg>` may nest within `<lg>`, i.e. stanzas within a verse passage.

Except in exceptional circumstances and following consultation with an editor, all verse passages (even if they are only one line long) should nest in `<lg>`.

<newtonSymbol/> [Non-standard character, i.e. anything other than the Roman alphabet, Roman or Arabic numerals, standard punctuation marks and anything covered by the Entity Set. Contains a succinct natural-language description of the character.]

Attributes:

****xmlns** (XML namespace). This is always "http://www.newtonproject.sussex.ac.uk/ns/nonTEI". However, transcribers do not need to worry about this value: it will be automatically inserted by your text editor.

** *value*: suggested values: "asterisk", "obelus" (note or `<addSpan/>` indicator looking a bit like a dagger), "cross", "2-barred obelus", "cross with 2 uprights", "hash character" (#), "dot in a circle", "circle surmounted by a cross", "cross surmounted by a circle", "dot in a circle surmounted by a cross", etc. If it gets more complicated than this (and it often does), leave it out and put in a `<!--TODO NP -->` tag - it will be up to the editors to come up with a consistent form of terminology.

Notes:

A certain amount of creativity on the transcriber's part is permissible in describing such symbols, but try to keep as closely as possible to the methodology outlined above. Above all, *be consistent*: don't refer to the same thing as "circle surmounted by a cross" in one place and "cross underneath a circle" in another. And *don't* try to use ASCII characters or special fonts.

If two or more asterisks (or whatever) function as a single symbol, treat them as such, e.g. `<newtonSymbol xmlns="http://www.newtonproject.sussex.ac.uk/ns/nonTEI" value="3 asterisks"/>`, rather than as separate symbols.

The standard punctuation marks that are OK are these:

. , ; : ? ! ' " () [] { } / \ ~

and '-', *if it's functioning as a 'hard' hyphen* (i.e. a hyphen that is actually intended to appear in the text, as in 'Baden-Powell' or 'Idol-Temples', as opposed to one that merely marks a line break in mid-word): if it functions as a dash, encode it as '‐'. Any other symbol not covered by the Entity Set should be encoded as `<newtonSymbol/>`, even if there's a character for it on your keyboard.

If faced with an alphabet other than the Roman, Greek or Hebrew (Greek and Hebrew are covered by the Entity Set), consult an editor.

<note> [Annotation featuring in the text, or editorial comment on the text.]

Attributes

* *place* (only required in transcriptions of manuscript notes): where the note physically appears (as distinct from its point of reference, which is indicated by the `<anchor/>`). Permitted values: "marginRight", "marginLeft", "top", "bottom", "p041v-lower" (further down f. 41v) etc., "p041v-higher" etc., or `xml:id` value of the page it's on if it's on a different page from its point of reference, plus, in the last case, any of the foregoing if it's not in the main body of the page. For instance, `<note place="p083r-marginLeft">` means the note appears in the left margin of f. 83r but its point of reference is not on f. 83r. If the note isn't all in one place, two or more *place* values can be entered, separated by single spaces, as for `<add>`. But `<note place="p083r-marginLeft">` has a hyphen rather than a space because both values apply simultaneously.

type: describes the type of note. The only acceptable value at present is "editorial" – used for a significant explanatory or codicological note by the editor. It will not normally be used by transcribers unless given prior clearance by an editor: transcriber's comments on the text should otherwise be confined to `<!-- -->` tags. No *type* value is needed for annotations that actually feature in the document.

* *target* (only required in transcriptions of manuscript notes): the `xml:id` value of the relevant `<anchor/>` (q.v.), prefixed by a hash character (#).

hand: code for the person who's added the note (see Name Codes). Only needed if the note is in a different hand from that of the main text.

resp: code for the editor responsible for the note. Only used in `<note type="editorial">`.

n (only used in transcriptions from print): the letter, numeral or symbol (encoded as an entity if it's a symbol) functioning as a note indicator in the original text. Not needed if there is no note indicator in the original.

Notes

`<note>`s should be transcribed at the point in the text to which they refer, and in the case of manuscript notes immediately after their `<anchor/>` (see the examples above under `<anchor/>`). If it isn't clear what a note refers to, put it where you think best and point the problem out in a `<!--TRANSC -->` tag.

In transcriptions from manuscript, any note indicators should themselves be transcribed, both where they appear in the text and where they appear (if they do) in the `<note>`s. In transcriptions from print, they should simply be recorded as the *n* value of `<note>`.

It is theoretically permissible to assign more than one `<note>` to a given `<anchor/>` (i.e. give two separate `<note>`s the same *target* value), though this has never yet arisen in practice. More importantly, Newton quite often has two or more points of reference for a single note (e.g. if he quotes the same work two or more times in rapid succession, or includes references to a single note in multiple drafts of a given passage). In such cases, give each `<anchor/>` a unique `xml:id` value (as explained above under `<anchor/>`), transcribe the note immediately after the first `<anchor/>`, and give the `<note>` two or more *target* values, separated by single spaces, corresponding to the `xml:id` values of the `<anchor/>`s that refer to it, e.g. `<note target="#n015r-01 #n015r-03 #n015r-04">`. (This may sound impossibly complicated, but will hopefully make sense if you encounter the situation in real life.)

If necessary, `<note>` can be divided into `<p>`s (paragraphs), even if this means that one `<p>` nests indirectly inside another, thus: '`<p> ... <anchor/><note><p> ... </p><p> ... </p></note> ... </p>`'. Normally, however, the text of the note goes directly inside `<note>`.

If necessary (as it occasionally but very rarely is), `<note>` may nest within `<note>`.

rend. Permitted values: "center", "right" (right-aligned) or "indent*", * being replaced by the approximate number of spaces by which the first line is indented (i.e. the number of letters that would comfortably fit into the indent), in multiples of 5 up to a maximum of 40. Normally, paragraphs are indented by about five spaces: in these cases no attribute is needed. More pronounced indentation is indicated by "indent10", "indent15", etc.; no indentation at all should have *rend* coded as "indent0". Not needed for minor variations in paragraph indentation, only for ones that really stand out and hence can be construed as semantically significant.

<pb/> [Page break.]

Attributes

* *xml:id* (only required for transcriptions from manuscript or early modern print): letter p followed immediately by the relevant page or folio number expressed as a three-digit string followed, if appropriate, by r (recto) or v (verso). Each **<pb/>** must have its own unique *xml:id* value within the document.

If the document isn't paginated or foliated, the best thing to do is to number the folios 001, 002, etc., and distinguish individual sides by following the number immediately with 'r' or 'v' to indicate recto and verso. Once you get past 9, drop the first leading zero: 010, 011, etc., and once past 99 drop leading zeros altogether: 100, 101, etc. Using microfilm images, however, it isn't always possible to be sure what is recto and what verso: if dealing with such a document, simply number the individual sides in the order they occur in and forget about 'r' and 'v' (and we'll sort it out when we reach the check-against-original stage).

If the document *is* paginated or foliated, follow the original numbering as far as possible, but if this involves giving two different pages the same *xml:id* value, e.g. if two folios have both been numbered 24, call the first one "p024r" and "p024v" as normal, and the next one "p024Ar" and "p024Av". If a folio number has got missed out, e.g. there *is* no f. 24, follow the original foliation but comment on the omission in a `<!-- CODIC -->` tag.

* *n*: the same number as the *xml:id* value but expressed in human-readable form with no preceding letter and no leading zeros, e.g. "24v" instead of "p024v". (This generates the page number that users will actually see on their screens; the *xml:id* value is purely for processing purposes.) In the case of modern print documents, the *n* value is simply the original page number (this may have to be supplied if for instance the original edition does not show page numbers on the first page of each chapter).

Notes

Some manuscripts have been given two or more completely unrelated sets of pagination or foliation: in these cases, when assigning *xml:id* values to **<pb/>**s, follow the one that seems most logical in terms of how the document is ordered now.

If **<pb/>** is preceded and/or followed by **<fw>**, there should be no gap between these tags. If a word is split by a page break, the **<pb/>** and

any associated `<fw>` should sit directly in the middle of the word, with no space either side. Otherwise there should be one space either side, unless the `<pb/>` falls between divisions, paragraphs, line groups or lines of verse, in which case the spacing doesn't really matter.

Hyphens on partial words (including catchwords) preceding page breaks should be indicated by `<lb type="hyphenated"/>`.

It sometimes happens, usually in `<addSpan/>` passages or `<note>`s, that the text moves onto the same page at two different points, each of which needs to be indicated by `<pb/>`. In such cases, append a hyphen and a lower case letter to each `<pb/>`'s `xml:id` value to distinguish them, e.g. `<pb xml:id="p354v-a"/>`, `<pb xml:id="p354v-b"/>`, etc. The `n` value, however, is simply "354v" in both cases.

In the more chaotic documents, particularly those that have been split up, reordered, repaginated and/or otherwise mucked about with at some point since 1727, assigning `xml:id` values to `<pb/>`s can be a very confusing business calling for an editorial decision. At the risk of repeating myself, **don't hesitate to ask for help if you aren't sure what to do.**

`<rdg>` [Alternative reading. Use for places where two or more alternative readings are placed one above the other, or to note variants between two versions of a text.]

The alternative `<rdg>`s must appear within `<app>` (q.v.), which serves to link the alternative readings, much as `<choice>` serves to link alternative renditions.

Transcribers who are asked to note variant readings between exemplars will be given individual instruction. The instructions below apply only to cases where the original document itself offers variant readings.

Attributes

* *place*: values as for the `place` values of `<add>`, though in almost all cases one `<rdg>` will be "inline" and the other(s) "supralinear" and/or "infralinear".

Notes

This usually happens in Biblical passages where Newton presumably feels the translation is open to question, or in passages of his own composition where he's keeping his options open about which of two synonyms to choose.

Use only for places where there is a clear intent *on the author's part* to present two alternative readings. If you are uncertain about a reading yourself, use `<unclear>` and note any plausible alternatives as variant `<unclear>`s in a `<choice>` tag (see `<unclear>` below for further detail).

Examples

```
'And at that time shall Michael stand up the great Prince which <app type="authorial"><rdg place="inline">standeth for</rdg><rdg place="supralinear">is set over</rdg></app> the children of thy People' (Keynes 2, citing Daniel 12:1).
```

'wishing for a <app type="authorial"><rdg place="inline">benediction</rdg><rdg place="supralinear">blessing</rdg></app> & amp; desiring to carry away the holy medicines of body & amp; mind' (Yahuda 15.7)

<reg> see <orig>

<sic> [This element is used in tandem with <corr> (correction) and both are contained within a <choice> element. <sic> contains faulty text (or what you judge to be faulty text) as it appears in the document. <corr> contains what [you think] the author or scribe meant to put. If there shouldn't be any text there at all, include an empty <corr/> element and the type value "noText". If the text should be deleted but isn't, give the <corr> element the type value "delText". If there are two or more plausible corrections, enter them as two or more <corr>s nesting in a <choice> tag which itself nests within the <choice> tag containing <sic>.

<corr> may additionally have a cert attribute that indicates the degree of certainty about the suggested correction on a scale of "high", "medium" or "low". Where two alternative <corr>s nest in <choice>, this can be used to give relative weighting to the proposed alternatives. The cert attribute is only required if there is any doubt about the required correction.

Notes

<corr> *must* be included even if it's blindingly obvious what it should be.

To restore text that is missing because of manuscript damage or authorial or scribal absent-mindedness, use <supplied>, not <sic> and <corr>.

<sic> should normally be applied to whole words, not just the bit of the word where the mistake occurs.

<sic> should be used sparingly, for things that really are obvious mistakes, not to modernise or standardise the original spelling or punctuation. If in doubt, leave the <sic> tag off but mention your doubts in a <!-- TRANSC --> tag.

Examples

'the killing of the <lb/><choice><sic>the</sic><corr type="noText"/></choice> witnesses' (Yahuda 1.3)

'in <choice><sic>explaing<lb/>ing</sic> <corr>explain<lb/>ing</corr></choice> how (according to Montanus) the son might suffer without the father ...' (Yahuda 15.7)

'the holy Ghost never reproves any nation under the notion of committing fornication beside the revolting Jews in the old <choice><sic>Testamen</sic> <corr>Testament</corr></choice> & amp; revolting Christians in the new' (Yahuda 1.3)

<space/> [Space deliberately left blank in the text.]

Attributes

* *dim*: dimension. Permitted values: "horizontal", "vertical".

* *extent* and *unit* are used in tandem to express the amount of blank space:

extent gives the numeric value and *unit* gives the unit of measurement ("chars", i.e. characters, for horizontal space and "lines" for vertical space). Note that the *unit* value is always plural, even if the *extent* value is only "1". The *extent* value doesn't need to be too precise, and may be given as "unclear" if the state of the copy you are working from makes it impossible to gauge accurately, or if it is simply unclear whether space has in fact been deliberately left blank. In these cases, no *unit* value is required.

cert can be used on a scale of "high", "medium" or "low" where there appears to be a lacuna in the text but it is not clear whether it is deliberate (e.g. in the case of what appears to be an incomplete interlinear or marginal insertion).

Notes

Not permitted between <p>s or between <head> and <p>: has to be included in them.

Not needed if a page has only been partially written on, e.g. if a section ends halfway down a page and the text resumes on the next page, or if a page features two <addSpan/> passages for insertion elsewhere at different points and there is a gap between them. Only use it for gaps that are there for a purpose, e.g. to make a heading stand out, to mark a break between subsections, or because Newton meant to add something later but never got round to it (which is easily the commonest reason for horizontal <space/>s).

DON'T use <space/> to indicate the indentation of a line of verse or the first line of a paragraph: this should be recorded in the *rend* value of <l> or <p>.

Examples

```
'For by what was shewed in Posit <space
  dim="horizontal" unit="chars" extent="3"/> they
were to rise after this head' (Yahuda 1.5)
```

```
'... kingdoms of the true religion, as the Visigothic
the Ostrogothic the Vandalic <add
  place="supralinear" indicator="yes">the
Burgundian the <space dim="horizontal"
  extent="unclear" cert="medium"/></add> &
for some time the Suevoian' (Yahuda 1.3)    '<head
  rend="center">An historical account of two
notable corruptions of Scripture, in a Letter
to a Friend.<space dim="vertical" unit="lines"
  extent="2"/></head>
```

```
<p>Sir</p>
```

```
<p>Since the discourses of some late writers ...' (New
  College Oxford 361(4))
```

<supplied> [Supplied text where the original is illegible or missing for some reason but is recoverable (even if only conjecturally) from the context or by reference to another source.]

Attributes:

* *reason*: as for <gap/>, or "omitted" if the author or scribe just missed something out.

cert: degree of certainty about the suggested restoration on a scale of "high", "medium" or "low". Only required if there is any doubt.

Notes:

In the case of text being supplied from another source, mention the source in an <!-- APP --> tag.

You may very well find that the manuscript you're transcribing serves in itself as a source of <supplied> text if it contains multiple drafts of the same passage. In this case simply reference the page from which you have taken the supplied text in an <!-- APP --> tag.

If a catchword is not repeated at the beginning of the following page, supply it, with the *reason* value "omitted".

Examples:

'Nor did Iulius Bishop of Rome know any thing of it
whe<supplied reason="damage">n</supplied><!--
CODIC page edge is frayed - jy --> <lb/>he
wrote in defence of Athanasius' (Keynes 10)

'it was made known to all Egypt & cannot any
longer <supplied reason="omitted">be</supplied>
concealed' (Clark Ms.)

'the two Olive trees & two Candlesticks standing
before the Go<supplied reason="copy"
>d</supplied><!-- APP supplied from Revelation
11:4 --> <lb/>of the earth' (Yahuda 15.7)

'et una <fw type="catch"
place="bottomRight">cum</fw><pb xml:id="p036r"
n="36r"/> <supplied
reason="omitted">cum</supplied> basibus et
capitellis 18 cub.' (Babson 434)

'he died at Constantinople <supplied
reason="damage">in a bog-house</supplied><!--
APP supplied from Keynes Ms. 10 f. 1r -->
miserably by &the; effusion of his bowels' (Clark
Ms.)

<unclear> [Uncertain or conjectural reading.]

Attributes:

* *reason*: why it's unclear. Permitted values: as for <gap/>, except that if text is difficult (rather than impossible) to read because it's deleted, the *reason* value should be "del" rather than "illgblDel".

* *cert*: level of certainty about the proposed reading on a scale of "high", "medium" or "low".

Note

If you can think of two or more plausible readings, nest `<unclear>` elements in a `<choice>` element, using the `cert` values to indicate which (if either/any) you think most convincing (see above under **<choice>** for examples).

Name Codes

The commonest hands that figure in Newtonian or Newton-related manuscripts are listed below with the codes that should be assigned to them and declared in the `<handList>` section of the `<teiHeader>`. It is entirely likely that other hands will have to be added to this list.

Thomas Burnet tb
Guillaume Cavelier gc
Catherine Conduitt cc
John Conduitt jc
William Derham wd
Fatio de Duillier fd
John Flamsteed jf
Bernard de Fontenelle bf
Hopton Haynes hh
Samuel Horsley sh
John Locke jl
Thomas Mason tm
Richard Mead rm
Francis Meheux fm
Humphrey Newton hn
Isaac Newton in
Thomas Pellet tp
Ferdinand de Saint-Urbain fsu
William Stukely ws
Owen Swiny os
John Wickins jw
Nicholas Wickins nw

Unidentified unknown

If your manuscript has two or more hands which are unidentified but clearly distinguishable from one another, call them `unknown1`, `unknown2`, etc.