Team MARKUP Quality Control Checklist

Key Points

1. Don't just look over this checklist--it contains key points from the GoogleDoc and schema, but not all information. be sure to read over the new additions to the GoogleDoc and the SGA schema as well before finishing your quality control work.
2. If you added questions to the GoogleDoc, did you go back and make changes to these areas after your question was answered?
3. Sometimes people answered questions, but their answers were incorrect (it happens). Please also read through the chart and make sure your comments don't have other comments correcting them.
4. Did you validate your work? For each of your XML files, go to Document > Validate > Validate while in Oxygen. Look at the bottom of the page for a list of errors the validator found. Your encoding is not correct until you can validate the file and see no errors in that list.
5. Did you push your work correctly and to the right place? If you can see your name and update text next to the file on this page, you're golden: https://github.com/umd-mith/sg-data/tree/master/data/eng738t/tei

Main Issues

1. Almost every tag should have a closing tag (e.g. <add place="superlinear">some text</add>). The milestone tag is one exception.
2. Do not use <p>.
3. Stuff that's in blue font in the transcription files is in Percy's hand!
4. For any symbol (ampersand, dash, plus mark, etc.), there's a proper encoding (see next item for some of these).
5. We're now using unicode for the ampersand (&), caret (^), and em dash (-). These codes are:
   a. ampersand = &
   b. caret = <metamark>‸</metamark>
   c. em dash = —
6. Put <milestone unit="paragraph"/> just before the closing </line> tag on the last line of a paragraph.
7. <line></line> for every line
8. <zone></zone> we should use different zones for main text, marginal comments, and page/chapter/volume numbers ("By convention, page numbering or similar kinds of notes occurring at the top of a page should be placed within their own zone, and this should be the

first zone on the page.") Follow this order of precedence when encoding zones: notes at top of page (page numbering, chapter/volume numbers) go first, then main text, then marginal comments, than numbers at bottom of page.

    a. Every piece of marginalia should have its own zone.
    b. Don't forget to add an attribute to your zone (e.g. <zone type="pagination">). You should always have something like <zone type="main">, not just plain <zone>.

9. <del> is for short deleted text (words and phrases), for example short strike-outs:
    a. <del rend="strikethrough"></del>
10. <delSpan> is for deletions longer than words and phrases
11. <add> is for any text added in around the main text such as
    a. <add place="superlinear"></add>
12. For places that are obscured, look at <unclear>, <supplied>, and <gap> (each indicates a different type of obscurity): [this page describes when to use what](#)
13. Use <mod resp="#pbs"> instead of <mod resp="PBS">
14. Do NOT use <mod type="subst">, <mod type="additions">, or <mod type="deletions"> at all (that is, do not group deletions/additions). If you have the rare case of an addition beginning at the end of one line and winding around to the next line, you should use plain <mod></mod> to encase it.
15. Breaking up words: We are, for now, going ahead and breaking up words when
    i. 1) a word stretches from the end of one line onto another:
        1. <line>Hey, here's an exa</line>
        2. <line>mple of a correct word extending over two lines!</line>
    ii. 2) a deletion or insertion occurs in the middle of a word:
        1. <line>I forgot the "e" in the word el<add place="superlinear">e</add>phant</line>
16. Use <del rend="strikethrough">, never overstrike

Specific Issues

Questions	The Law
1. When a word is rendered wrong in the manuscript, should we find some way to indicate (**sic**)--for example, when Mary writes the pronoun "I" as "i" in 0002.xml line 15? I feel we want to retain the lowercase here, but I'd also be interested in indicating it isn't just a transcription error on my part (AmandaV)	We're not using <sic>.Leave typos and grammatical errors alone-they'll be changed later.
2. In the particular case I have (0012.xml line 15, Clifford) she writes "expreped" which I assume means "expressed;" **should I use <choice> as well as <sic>** to include my assumption? Thus far I've just been treating it as <sic> because it seems a bit of a leap to	Don't use sic. If unclear, use <unclear> and indicate your level of certainty about the guess you provide

read "expreped" as "expressed," yet it is the only thing that makes sense in the context and with the letters provided.	

3. I have a case (0012.xml line 18 and 19) where a **word is missing**. The sentence reads "The followed and they disappeared [....]" Now, I can tell from the
context that it ought to read, "The girl followed" so how should I mark it? As a choice (though I'm suggesting a correction that can be substituted for anything in the text, but rather adding to it)?
	For places that are obscured, look at <unclear>, <supplied>, and <gap> (each indicates a different type of obscurity) E.g. group removed obscurity with replacement:
	<subst>
	    <del>
	    <gap reason="illegible" quantity="5" unit="character"/>
	    </del>
	    <add>
	        apple
	    </add>
	</subst>

4. I'm seeing a lot of **symbols** that function as ampersands, but that visually look more like strange uppercase As. Should I simply render these as ampersands? (AmandaV; e.g. 0002.xml line 17 where there's a symbol that semantically should be an ampersand but doesn't quite look like one)	We're now using unicode for the ampersand (&), caret (^), and em dash (-). These codes are:
	ampersand = &#x0026;
	caret = <metamark>&#x2038;</metamark>
	em dash = &#x2014;

5. How do we treat **circled words?** (AmandaV)	Use rend=bordered
6. Relatedly, how do we indicate a **circled page number**? (AmandaV, 002.xml, top-left)	Like this: <zone type="pagination" **rend="bordered"**>4</zone>
7. Should we distinguish a "**write-over**" (line 8: a lowercase "i" is written over the capital "I" in "In") from an **overstrike**, and if so, how? (Jen)	Use "strikethrough" for all deletions, no "hashmarks" or "doublestrikes", etc. In many cases, we would be providing more information than is actually needed. There is no need to be this specific.
8. How have people handled **corrections written into the manuscript that are not above or below lines**. I have the word 'him" with the M obviously added in later because Mary wrote too fast and only put "hi" first (that's my baseless assumption), but the addition fits into the line. I'm thinking it should be <mod="????"> (Nigel).	Use <add place="intralinear">
9. Do we need to **tag people, places, and organizations** with <name>? (AmandaV) What about numbers?	No. We aren't doing this for the main project but this is easily an area where all of you might choose to add more encoding if you do your

	final project around markup.
10. How do we indicate **double strike-outs** (0002.xml line 20 uses two lines to cross out a single character; AmandaV)	Use "strikethrough" for all deletions, no "hashmarks" or "doublestrikes", etc. In many cases, we would be providing more information than is actually needed. There is no need to be this specific.
11. While we are on the subject, actually, **do we want to differentiate between strokes when crossing something out.** For example (0012.xml line 19), the word "wiping" is crossed out, there is a break in the overstrike and then "dry (unknown word) her tears whi" is crossed out. Is it important to mark them as separate deletions? I at least have been treating them as such (Clifford).	Use "strikethrough" for all deletions, no "hashmarks" or "doublestrikes", etc. In many cases, we would be providing more information than is actually needed. There is no need to be this specific.
12. What do we do with **pencil** (at least I'm assuming that is what they are) marks on the page (0012.xml and 0013.xml, Clifford)? Occasionally I can make out words, but in most cases the pencil marks are overwritten in ink and too indistinct to make out.	I think if there is pencil inline that has been overwritten that you can't make out, you might encode it as an add and a del, e.g. <mod><add place=intralinear>current visible text><del rend="overwritten pencil"><unclear>Text you can't make out</unclear></del></mod> For that marginal intervention in pencil (in 0012), I would suggest adding another <zone type="left_margin"><unclear>…</unclear></zone>
12a. In pencil, but a different sort of problem, do we wish to mark **doodles** in the margins of pages (0013.xml, Clifford)? I have on that page what appears to be a small bundle of flowers drawn in, but I can't think of how to encode it. While perhaps not pertinent to the text, it is certainly interesting and delightful. I have some doodles from Percy as well (0003.xml). I'd love if we could encode these, if only because it's interesting that Percy is doodling on Mary's manuscript (what does that say about his editing process and attitude toward the text?) AmandaV	Yes, we want to capture this. I would suggest including a <figure><desc>PBS doodle or sketch</desc></figure> inside the left margin zone. If you are feeling ambitious, you might add @type to figure and select a best-fit value from iconclass: http://www.iconclass.org/help/outline
13. How do we want to indicate **underlining** in the text (0020.xml lines 21 and 22, Clifford)?	<seg rend="underline">…</seg>
14. Should we note when the **handwriting is somehow messy or strange** in some way? My example is the word "retreat" angled slightly at the end of the page in order to make it fit (and prevent the word from	No, this is more specific than we need.

16. (0022.xml, Jen) -- Is there a **preferred order in which to render supra and sublinear** mods (spatial > logical)? For example, the logical progression here is sub to supra ("snowy ground"), so I'm inclined to indicate the mod in that order...	Insertions and deletions should be next to one another, but there is no real consideration for order. (Don't forget we've moved from supralinear to superlinear)
17. The transcription files give a lot of information about the physical page (e.g. ink blots, burns), as well as hand information. Should we be encoding all that information as well, or focus on encoding just what we can get out of the image files without the advanced knowledge of the **transcription's footnotes?** (AmandaV)	Trevor and Neil verified we don't need to include *all* the info in the footnotes of the transcription files, BUT we should include it if it's important to understanding the encoded manuscript. We should use our judgement and encode anything in the footnotes that does seem important (e.g. always check them to see if Percy's hand appears anywhere on the page); but since the transcription notes were meant to exhaustively replicate the actual manuscript pages, we do not need to encode at the same level of detail as represented by the footnotes. Remember that Percy's writing is given in blue text and always had <mod rend="PBS></mod> for his writing.
18. How to use the **anchor tag** to connect marginalia with the line(s) it's next to?	This page has a good example (use ptr within the main text, then use anchor in the left-margin zone). See also 0002.xml for an example of two pieces of marginalia encoded in two left-margin zones via anchors. Anchors are used when a deletion (or something else) crosses multiple lines. You can either treat it as multiple deletions (or other tags) or use an anchor. Example: <line> the <delSpan SpanTo="#abc"/> tree </line> <line> was <anchor="#abc"/></line> No x,y coordinates are being used to denote the spatial area of marginal notes. Just use simple "left_margin" or "main"
19. Anyone know why the "**surface" tag** here is giving me trouble? It looks identical to the test-page we did in class during Bootcamp (which is valid) and to AmandaV.'s .0002.xml but mine seems to be asking for "partOf" something. The <surface> is closed with a </surface>, by the way. This problem is all that exists between me and pushing the mark-up. (MichaelG, .0032.xml)	<surface> now requires the attribute partOf. You'll need to type in partOf="ox-ms_abinger_c57" right before the final > of your surface tag (near the beginning of each of your XML files). AmandaV
20. How do we tell the difference between **page numbers in original**	The original pagination is what is circled and the library pagination is

hand and library (i.e. <zone type="pagination">) numbers? And a<u>re library numbers assumed to be circled</u> (i.e. do we need to encode the circling)?	not. (See below chart for an image that shows a good example*). I would encode this as: <zone type="pagination" rend="bordered">4</zone><zone type="library">3</zone>
21. How are we using **delSpan**?	Use delSpan for long deletions (more than a word or phrase). To indicate a deletion spanning lines & referencing an anchor tag signifying the end of the deletion; this anchor tag is entered at the endpoint of the deleted segment. E.g. <line>the<delspan spanTo="abc"/></line> <line>was<anchor xml:id="abc"/></line> addSpan works similarly for insertions
21b. Are there specific terms we should use when **creating anchor ids**?	I think it is okay to use any number for the anchor ID, as long as it is unique on that XML page (i.e. if you need to have multiple anchors/ IDs, use a different number for each one). I followed a pattern that I saw other SGA files doing: c57-0002.01 (file number plus a .0x as the unique part). Make sure you use <anchor xml:id="abc"/>
22. Is there any special way that we are marking **vertical deletions** (e.g., [page 79](#))?	No, just use <del rend="strikethrough">
23. When we want to put in a **<zone type="left margin">**, do we insert it where it falls in the image? That means closing and reopening the <zone type="main">. Or do we just insert a <zone type="left margin"> after <zone type="main">? Will this display correctly? We might also try using a ptr target, but how will this display of there are several lines of marginalia? ([Page 79](#) once again provides an example.)	See notes on <anchor> and <zone> usage above. Zones have a specific order that is *not* how they fall on the page. Follow this order of precedence when encoding zones: notes at top of page (page numbering, chapter/volume numbers) go first, then main text, then marginal comments, than numbers at bottom of page. When you've got marginalia, ptr should be in the main text to indicate there is marginalia next to that line, and then when you get to your left-margin zones at the bottom you can use anchors to link the marginalia to those ptr. 0002.xml has two examples (don't forget to create one zone per piece of marginalia)
24. Is there a way to indicate "**subscript**"? I've got the indefinite article "a" switched to "an" & the addition is not truly sublinear. Maybe I'm not navigating the Encoding Guidelines correctly, but I'm not finding instructions for this in there (Phil).	Just use sublinear. Subscripted insertions are treated as sublinear in schema.

25. Same for the **em-dash**! I'm inserting a double-dash plus a note for now. Not satisfying somehow. (Phil)	We're now using unicode for the ampersand (&), caret (^), and em dash (-). These codes are: ampersand = & caret = <metamark>‸</metamark> em dash = — Be careful just doing a blanket find&replace for these, since there is another type of hyphen used a lot in the header info
26. Are we using **<p> or <milestone>** to marks ends of paragraphs? And where (if anywhere) does <milestone> get used?	Don't use <p> at all. Use </milestone> to separate an old and a new paragraph by putting <milestone unit="paragraph"/> just before the closing </line> tag
27. I had everything but "partOf" validated on Monday, but now, after fixing "partOf," Oxygen shows my rend="**supralinear**" as problematic. Did the schema change? Am I overlooking something? (MichaelG)	Use superlinear instead of supralinear. Also, mods with additions now call for you to put place="superlinear" rather than rend="superlinear".
28. How to render **^** mark?	^ should be considered metamarks, encoded like this: TEXT <metamark>‸</metamark><add place="superlinear">the added text...etc. Do not treat them as lexical additions, i.e. don't just type ^ alone.
29. How do we account for **lines** drawn across the page to suggest a new page in the text? I tentatively have <milestone unit="page"/>. Additionally, Is this still part of the 'main' zone? (0033.jpg, MichaelG)	<metamark rend="horizontal line"></metamark>
30. If a metamark (^) occurs below the middle of a word, should we encode it in the midst of the word, or either before/after it? For example, 0008.xml line 29 gets an e added into the middle of a word. AmandaV	Breaking up words: We are, for now, going ahead and breaking up words when a) a word stretches from the end of one line onto another: <line>Hey, here's an exa</line> <line>mple of a correct word extending over two lines!</line> b) A deletion or insertion occurs in the middle of a word: <line>I forgot the "e" in the word el<add place="superlinear">e</add>phant</line>

31. Do we want to do anything special with *half-sheets* of paper? (see: c57-0035.jpg; MichaelG)	No. Just encode the lines you see.
31. ...or sheets completely blank, except for an ink blotch? (see: c57-0036.jpg; MichaelG)	Yes. Input <desc>Blank page</desc> *before* the graphic tag in the header, as here: \<surface xmlns="http://www.tei-c.org/ns/1.0" xmlns:sga="http://sga.mith.org/ns/1.0" xml:id="ox-ms_abinger_c57-0011" ulx="0" uly="0" lrx="5368" lry="7104" partOf="ox-ms_abinger_c57"> **<desc>Blank page</desc>** <graphic url="../../images/ox/ox-ms_abinger_c57-0011.tif"/>
32. WRT <mod> & its deprecation except where resp="#pbs" applies: What is the structure now? As quoted from above: "stop using and remove all use of the <mod> tag, except for <mod resp="#pbs"> and in the rare case when an addition begins on one line and runs into the next line or left-margin. Just use plain <mod> for this; don't use <mod type=?> any more." So we are not using <mod type="subst">~~something crossed out~~<add resp="#pbs">Percy's addition</add></mod> now? Or are we? (Phil)	No <mod>. Unless for a change that spans more than one line (addition runs across two lines e.g.) -- then just use <mod> w/out type descriptor. So, your example should read (note that I added a few things--your adds should always have a place, and your dels should always have a type): <del type="strikethrough">something crossed out</del><add place="superlinear or sublinear"><mod resp="#pbs">Percy's addition</mod></add>
33. What about deletions that are smudged-over / undeleted? Mark the undeletion as an adjacent addition? (i.e. show "use" where it is stricken-through as a "del" & then show "use" where the strike-through is smudged out as an "add"; or, alternatively, embed the "del" within an "add" to signal the cancellation of the "del"?). Embedding seems to me to capture the logical structure pretty naturally; so, the outermost layer is the last action that occurred. (So, a superlinear addition is made; it is placed adjacent to what it replaces, or (if within a word) at its insertion point; & if this superlinear addition	Your proposed solution works fine. Insertions and deletions should be next to one another, but there is no real consideration for order.

is then stricken out -- we figure it has to be there before it is stricken -- then embedding it in a <del> tag would signify it is deleted after addition. So, proposed: ~~prior intralinear text~~<del rend="strikethrough"><add>superlinear addition that ends up stricken-through</add></del> continuation of intralinear MS encoding here (Phil)	
34. How do we know what is in Percy's hand?	It's in blue in the transcription!
35. How is unicode being parsed--is 0026 okay, or should we use just 26?	0026 is fine.
36. Is there any sort of modSpan capability to indicate that a block of text is in PBS's hand? For an example, see the marginal comments on page 78.	I think you'd use <addSpan place="wherever"><mod resp="#pbs">Block of text in Percy's hand</mod></add>
37. I have a lot of math that Percy and Mary are working out in the margins (c57-0037.jpg). I've anchored the text, but I don't know how we want to code the math since addition/subtraction moves vertically down the page but, in coding, might move horizontally (e.g. 18+18=36). (MichaelG)	Trevor needs to figure this out, so he said that the placeholders you have right now are fine for our assignment. No need to change anything. <zone type="left_margin"> <line>MATH<anchor xml:id="c57-0037.02"/></line> </zone>
38. Sometimes one of the Shelleys places an "X" in the text to gesture to a note in the margins (also indicated with a corresponding "X"; see c57-0041.jpg, for example). Is there a mark we should make in lieu of just writing a capital X? (MichaelG)	<figure><desc>PBS insert mark X</desc></figure> (If it's Percy's--the transcription footnotes will say, but it won't be in blue even if it is his--don't forget to surround with the mod tag)
39. Another delSpan question (to be merged w/ above and canceled here when answered):	You're doing it right. To indicate a deletion spanning lines & referencing an anchor tag signifying the end of the deletion; this

I got an Oxygen error flag telling me various things were disallowed when I tried to close my delSpan with an anchor tag. I was finally able to eliminate the error message by writing something appearing to conform to the Encoding Guidelines as follows: <anchor xml:id="nextline"/> The difference between this and ~~the proper form given in answer to (18) above.~~ ~~<anchor="#abc"/>~~ ~~...is obviously the "xml:id".~~ So... I wonder if I'm doing it right using "xml:id" (Phil)	anchor tag is entered at the endpoint of the deleted segment. E.g. <line>the<delspan spanTo="abc"/></line> <line>was<anchor xml:id="abc"/></line>
40. What if a page is missing in the reference images? (There is a missing page between images c57-0086 and c57-0087)This sounds like a general project management question, but it seems like it's worth consulting someone about it.	For our purposes, we don't need to think about this. The pages are looseleaf and often got shuffled around or lost; the SGA staff editors will handle this.
40. How do we indicate a new paragraph has been started?	Implicitly by previous paragraph having ended. Before closing the line before a paragraph ends with </line>, write <milestone unit="paragraph"/>
41. What do we do when a mark has simply been erased?	Try: <del type="erased"> the erased text </del> Surround the erased text with <unclear reason="illegible"></unclear> if it isn't clear to read. If you can't read the erasure at all, try: <del> <gap reason="erased" quantity="5" unit="character"/> </del> (where 5 is the number of letters you're guessing were erased)
42. I've found a circling of multiple words that appears in fact to be a	You can add a rend="bordered" to indicate any type of circling. I'm not

multi-line transposition mark -- which is NOT apparently applied in the production of the Frankenstein full text I just looked at -- so I think it may be an intended emendation that is lost in publication. (On top of this, a careful look at the result shows it is slightly illogical -- but it's obvious that the resulting change, fixed slightly, works better than the apparently published text.) So--I see how to do a transpose in the Encoding Guidelines. But how to mark the blocks of text as "circled" (really, partially circled, with transposition mark)? I haven't quite got that together. Ideas?	sure what you mean by transposition mark--could you describe or add a screenshot?
43. Are we doing anything to record when the ink is smudged (but words are still readable)?	<unclear reason="illegible"></unclear> (you might even be able to substitute "smudged" for "illegible"; try and see if this validates)
44. How are we noting the edits that are Percy's?	surround them with <mod resp="#pbs"></mod> (this is in the notes above this chart--please be certain you're reading through those)

* image removed because of copyright