

ENDINGS: USING DIAGNOSTICS

Joey Takeda

Junior Programmer

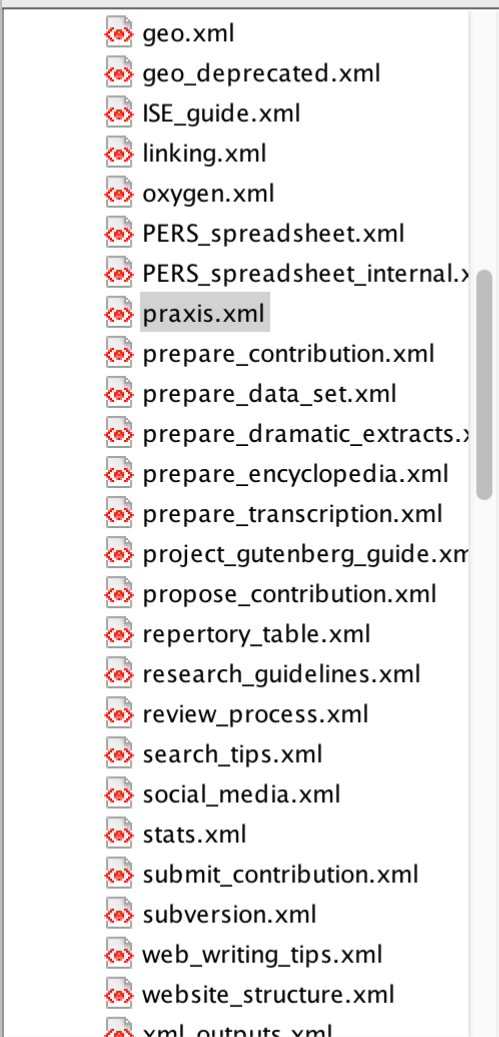
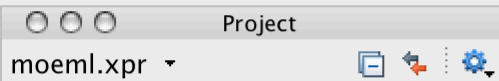
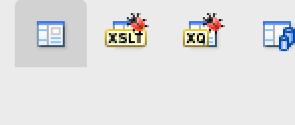
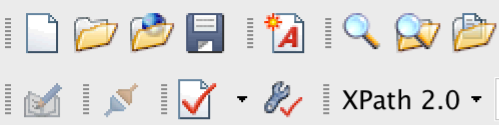
The Map of Early Modern London

WHAT ARE DIAGNOSTICS

- A generated file (usually HTML or text) that displays common encoding errors and problems that are not usually caught by schema/schematron
- Validation catches for local errors in the data encoded in an original file and does not usually analyze linked data
- Diagnostics interrogate the linked data between documents, while also checking local errors, issues of consistency, and potential errors

WHAT VALIDATION CAN DO

- Check that all ids are in the proper form (eg: if the file is named “BLOG2” then any ids created further in the file must start with “BLOG2”)
- Ensure that elements of the file are included before particular stages of the document (eg: All published location files **must** include an abstract)
- Enforce stylistic conformity (eg: Straight apostrophes are not allowed, use curly apostrophes)
- +++++++



```

369          <item><ref target="mol:editorial_style#editorial_style_parenthetical">Parenthetical
369 Citation</ref></item>
370          <item><ref target="mol:editorial_style#editorial_style_workscited">Works Cited
370 List</ref></item>
371        </list>
372      </item>
373    </list>
374  </cell>
375
376  <cell role="data" rows="1" cols="1">
377    <list type="bulleted">
378      <head>Social Media Guidelines</head>
379      <item><ref target="mol:social_media#social_media_culture"><title level="m">MoEML</title>'s Social
379 Media Culture</ref></item>
380      <item><ref target="mol:social_media#social_media_etiquette"><title level="m">MoEML</title>
380 Etiquette</ref></item>
381      <item><ref target="mol:social_media#social_media_frequency_style">Frequency and Style of
381 Posts/Tweets</ref></item>
382      <item><ref target="mol:social_media#social_media_formatting">Formatting Tweets</ref></item>
383      <item><ref target="mol:social_media#social_media_blog">The <title level="m">MoEML</title>
383 Blog</ref></item>
384      <item><ref target="mol:social_media#social_media_resources">Related Resources</ref></item>
385    </list>
386  </cell>
387 </row>
388 </table>
389 </div>
390
391 <div xml:id="praxis_further">
392   <head>Further Resources</head>
393   <table rows="1" cols="2">
394     <tr>

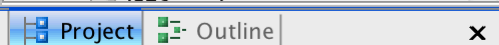
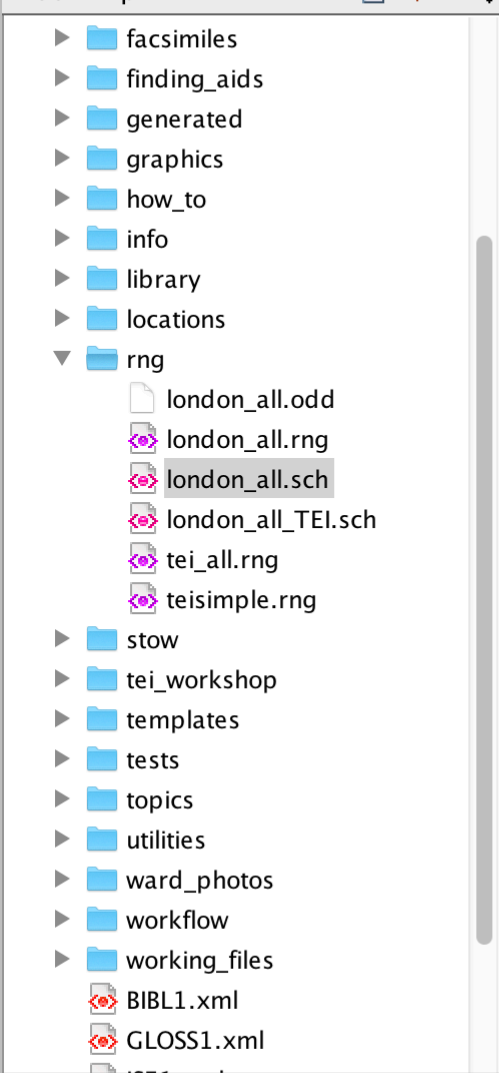
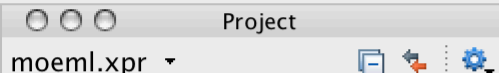
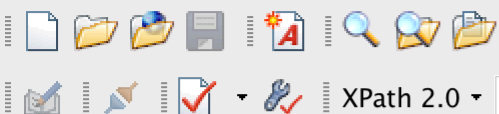
```

🔍 E [ISO Schematron] ERROR: "Straight apostrophe" characters are not permitted. Please use the Right Single Quotation Mark (U+2019) character instead.

Text	Grid	Author
------	------	--------

WHAT IT CAN'T (OR SHOULDN'T)

- Check to see if an @xml:id is unique out of 10742 other ids across over 1600 files
- Test to see if an a particular pointer actually points to something in another file
- Double check if the thing you're referencing is actually what you're claiming it is (eg: tagging a person as a location)
- Give a suggestion; even though you can differentiate between warnings and errors in schematron, the ominous line still makes you think what you're doing might be wrong even if it isn't



● stow_1598.xml ✕ ● london_all.sch ✕

```

388      <pb facs="moleebo:18464|2"/>
389      <!-- printer's ornament -->
390      <div type="dedication" xml:id="stow_1598_dedication" n="Dedicatory Epistle" change="#stow_1598_dedication_status">
391          <!-- We've included the blank verso of the title page in the Dedication div for now. -->
392          <head>TO THE RIGHT Honorable, the Lord Mayor of the <ref target="mol:LOND5">Citie of London</ref>, to the communalitie,
393              and <hi>Citizens of the fame, <name ref="mol:STOW6">Iohn Stow</name> Citizen, wifheth long health and
394              felicitie</hi>.</head>
395
396          <p>Since the first publifhing of the perambulation of <hi>Kent</hi>, by that learned
397              Gentleman M. <hi><name ref="mol:JOHN8">William Lambert</name></hi> Efquire, I haue hearde of fundry other able
398              perfons to haue (according to the defire of that au<lb rend="hidden" type="hyphenInWord"/>thor) affayed to doe
399              fomewhat for the particular Shires and
400              Connties<note type="marginal" resp="mol:MILL2">I.e. Counties.</note>, where they were borne, or dwelt, of which
401              none that I knowe (fauing M.
402              <hi><name ref="mol:NORD2">Nor<lb rend="hidden" type="hyphenInWord"/>den</name></hi>, for the Counties of
403              Middlefex, &amp; <hi>Hertford</hi>) haue vouchsafed their labors
404              <fw type="catchword" rendition="#stow_1598_catchword">to</fw>
405              <fw type="signature" rendition="#stow_1598_signature">A2</fw>
406              <pb facs="moleebo:18464|3"/>
407              <fw type="header" rendition="#stow_1598_runningheader">The Epiftle</fw> to the commo good in that behalf. And
408              therefore
409              concurring with the first, in the fame defire to haue drawn together fuch fpeciall
410              defcriptions of each place, as might not onely make vp an whole body of the Englifh
411              Chronographie a<lb rend="hidden" type="hyphenInWord"/>mongft our felues: but alfo
412              might giue occafion, and courage to M. <hi><name ref="mol:CAMD2">Camdin</name></hi> to increafe and beutifie his
413              fingular worke of the whole, to the view of the learned that be abroad. I haue
414              attempt<lb rend="hidden" type="hyphenInWord"/>ted the difcouery of <hi><ref target="mol:LOND5">London</ref></hi>,
415              my natie foile and Countrie, at the defire and perfwafion of fome my good friendes,
416              afwell becaufe I haue feene fundrie an<lb rend="hidden" type="hyphenInWord"/>tiquities my felfe touching that
417              place, as alfo for that through fearch of Recordes
418              to other purpofes, dyuers written helps are come to my handes, which few o<lb rend="hidden"

```

⚠ W [ISO Schematron] WARNING: The @type='marginal' value on the note element is deprecated in MoEML documents. You should convert these to use the @

Text Grid Author

MOEML'S DIAGNOSTICS

- Diagnostics are generated as part of the regular build process
- Checks the encoding of the data, especially in relation to the data points to which the encoding links
- Checks for possible or “fuzzy” errors in the data: things that a computer can flag up as potentially wrong, but requires human intervention

[SCREENSHOT OF MOEML'S DIAGNOSTICS]

ENCODING ERRORS

- Duplicate ids (breaks the build completely)
- <name> instead of <ref>
- Organizations tagged as <name>s but without an @type='org'
- Bibliography entries tagged without @type='bibl'
- Malformed ids (mostly typos)

BENEFITS

- While validation scenarios could potentially check for some errors like duplicate ids or malformed ids, it would require significant processing energy
- Catches typos and forces editors to go back and review their work, thus operates as a form of proofing
- Gives a graphic representation on how “healthy” the data is
- An easy way for large projects to chip away at legacy code
 - Every team member takes 10 minutes out at the end of the day to fix a few errors

“FUZZY” ERRORS

- These are issues that aren't technically problems, but would make the site better if someone would fix or address them

[SCREENSHOT OF MOEMLS DIAG]

BENEFITS

- Checks for possible errors without making something that is completely valid erroneously invalid
- Allows for decisions about project practices to be made and implemented as time allows, especially for big changes that would cause the majority of the project to become invalid
- **Problem:** How do we declare in our encoding that something that might look incorrect is actually correct?

CASE STUDY: STOW_1598

- Alongside our implementation of the regular site diagnostics, we have also created a similar system for our process of proofing the 1598 edition of Stow
- Proofing a text of this size is a challenge, especially since it requires that we have good definitions of what it means for something to be “finished” on the site
- Before we send chapters of Stow to peer-review (declaring that it is “finished”), we must ensure that everything that is included is also finished

Stow 1598 Diagnostics: Cornhill Ward

Statistics ▼

TEI documents found:	1552
<bibl> entries found:	1470
<person> entries found:	3255
<org> entries found:	103
<ref>s pointing to tagged toponyms found:	141
<name>s pointing at people (in PERS1.xml) found:	67
<name>s pointing at organizations (in ORGS1.xml) found:	6

Incomplete locations (31) ►

Incomplete personography entries (53) ►

Incomplete orgography entries (4) ►

<ref> used instead of <name> (0) ►

<name> used instead of <ref> (0) ►

CONCLUSION

- Diagnostics:
 - Allow encoders to check their work and make sure everything will function the way they want it to
 - Maintain site consistency and usability
 - Give a rendering of data's readiness for proofing/publishing purposes
- However, diagnostics can be misleading if there is no standardized way of declaring that an item has been checked and is certainly not an error