

2014-08-27 Status Report

Summary Points

- A solid first phase conversion program is complete.
- Text for years 1767, 1768, 1769 converted to TEI, validated, and loaded to PubMan. Estimate for balance of years complete.
- Notes for both volumes converted to TEI, validated.
- Issues and future work documented
- 42.55h spent to date.

The ongoing data work described here is now superceded by the [Road Map](#).

State of the Data

The first phase of migrating the MB files from the old Scribe and TeX typesetting files into TEI was focused predominantly on building as much structure from the source as possible. The context provided by the XML structures is a better basis to begin further enrichment.

Notes

The notes have been organized into textual divisions by year, assigned unique identifiers, and the old note numbers retained for both mapping to their references and preservation. Some initial efforts to mark titles within them were made, and font changes preserved that will be modified to semantic markup where possible.

Text

Using the print copies of the Memorandum Books as a guide, the entries are laid out in tabular fashion though not slavishly so since the digital medium offers some flexibility print did not. The basic layout of entries is as follows:

```

# Columns
#   1      2      3      4
#   6.25%  6.25%  6.25%
# Year
# Month    Day      Entry for the day with
#                      a hanging indent.
#           Day    X   Entry for the day with an X, always includes a
date. Has a
#                      hanging indent.
#                      Entry for the same day, sans date or 'X' goes here
left aligned on entry column
#                      X   An 'X day' looks like this, left aligned on entry
column. Never has a date.
#                      X   A 'day list' indents to align left with hanging
indent above, always has an 'X' or a checkmark
#                      An 'X list' indents to align left with hanging
indent above. No preceding character.
#                      Sublists indent another em and have a hanging
indent
#                      Subsublists indent another em and have a
hanging indent (by default, usu. these indicate a special alignment)

```

Of course, this IS Jefferson, and the MB were created over a long period of time, so there are many permutations on the above to be found. There are also embedded tables within the text of list entries, and entirely separate charts/tables for things such as weather, etc.

Basic cleanup will consist of verifying the converted content against the print copies, though as previously noted there are differences in the print copy vs. digital source to be found. There is no need for the reviewers to worry about embedded tables, which for the moment appear outside of the entries they belong to (directly following them). They can be moved into their parent entries programmatically.

What is marked/handled

- Data is 100% TEI conformant.
- Font changes and languages (Greek)
- Editorial Conventions (supplied, deleted, etc.)
- Formatting instructions (as for interventions)
- (In the notes) some bibliographic references
- Some dates (just with <date/>, no ISO values generated yet)

What isn't marked/handled

- Embedded tables need to be moved into the list entries they are a part of
- Decide on strategy for notes (embedded or at end of file), implement and link to references
- Expand and improve date markup in both text and notes
- Currency
- Measures
- Court cases
- Abbreviations
- Links to Jefferson documents
- Names, Places, Organizations, Events
- Countries, other geographic locations
- (In notes) Bibliographic references and source links

Other Possibilities

- Given supporting markup (names, dates, amounts), generating transactionography records.

Remaining Work

- Convert the remaining years 1770-1826, validate, load to PubMan. Est. time to complete: 16-26h.
- Transliterate Greek in Notes (done in Text, only one note contains it) (1h)
- Discuss above points, prioritize, enrich data accordingly, update work plan

Notes About the Work Environment

- Editorial conventions don't quite follow the PTJ:RS use of `<supplied>`. In the MB, there may be more than one word with a low degree of certainty, so the question marks are retained in the data rather than resolved by the display rules. Right now this results in a double question mark being displayed in the document previews at the end of the insertion, e.g. `[supplied? also supplied??]`.
- We are using `<delspan/>` to mark non-spans of restored content. There are no style rules for this in the PTJ:RS PubMan displays so the tags will display in red. (Don't worry!)

Actions

- Make updates to PTJ:RS PubMan xsl stylesheets for editorial conventions so they display properly
- After review, meet and scope ongoing work