INTRODUCTION TO THE TEXT ENCODING INITIATIVE (TEI) FOR HISTORICAL DOCUMENTS

```
SIGNINGS CONTROL LONGER LAND IN SELE VERY LINE SELECTION OF CONTROL OF CONTRO
                  disposage of finiting int the Larger used. In Sets versions, and spartfully to contain a song or poss that seems distinct from the surrounding test and may not be obtained.
               tightings of "Torright"-Morks mon-English words. The -uttled lings out; uttribute identifies the foreign language, where openpriote. Jinglings-
             staplings II- now "Marks a proper now that may be quoted or limiticized in the text. -/ taplings-
            complete - tills Atres o title that may be quoted or itslicized in the text /tagliage-
         tagillage of motion Abrits areas where content in a prior source text is not present in the current reading of authors.
       AugUsage 11-mp "Critical apparatus containing variant resultings from prior publications. /tagbage-
 pr-Words, spaces, and punctuation characters are numbered supportially, incremental by 10.0%
guillast other elements begin with an element-specific prefix, followed by a reference to the Folgar Through Line Mamber, a sequential numbering of the numbered Lines in the to
```

Caitlin Pollock, @stilettobrarian HILT 2017, Austin, Texas

Background

- → Machine readable texts example: MARC (library) records
- → Text Encoding for the Humanities
 - **•** 1949
 - ♦ IBM and Father Roberto Busa*
 - ♦ Busa began using computing to perform literary analysis on the works of Thomas Aquinas
 - ◆ Searching for the concordance of words in Aquinas text corpus
 - ◆ Index Thomisticus was printed, but later released on CD-ROM and the web.
- → TEI traces back to Vassar College and the ACH.
- → Now a large research community or consortium

*See: Gold, Matthew K. (2012)"The Digital Humanities Movement." In Matthew K. Gold (Ed.), Debates in Digital Humanities. http://dhdebates.gc.cuny.edu/debates

Sooo... What is TEI?

What is TEI-C and TEI?

From the website: The Text Encoding Initiative Consortium is an internal organization whose mission is to develop and maintain guidelines for the digital encoding of literary and linguistic texts. The Consortium publishes the Text Encoding Initiative Guidelines for Electronic Text Encoding and Interchange: an international and interdisciplinary standard that is widely used by libraries, museums, publishers, and individual scholars to represent all kinds of textual material for online research and teaching.

Shorter: TEI is a standard for encoding text using an XML format.

The TEI Guidelines

The TEI Guidelines in English can be located here:

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ (and there is a copy in the GitHub repo.)

Two most important parts of TEI:

Text Body

Element References

The TEI Header

File Description <fileDesc>

- ◆ *Title Statement < title State > *
- ◆ Edition Statement <editionStmt>
- Extent <extent>
- *Publication Statement < publicationStmt>
- ◆ Series Statement <seriesStmt>
- ◆ Notes Statement <notesStmt>
- ◆ *Source Description < sourceDesc>

Encoding Description < encoding Desc>

Profile Description <profileDesc>

Revision Description < revisionDesc>