

## Metadata README for VEP Early Modern 1080

This corpus was generated for Serendip+Ity, a topic model visualization system (<http://vep.cs.wisc.edu/serendip/>). The corpus contains 1,080 plain text files, generated from Text Creation Partnership corpora. The texts were originally published between 1530 and 1799. The corpus was built by randomly sampling 40 texts per decade from a larger archive in an attempt to provide a less biased cross-section than just using well-known texts.

This metadata was prepared by Mattie Burkert (<http://mattieburkert.com/>) and Katie Lanning (<http://www.katielanning.com/>) as part of the Mellon-funded Visualizing English Print project at The University of Wisconsin, Madison, with supervision by Michael Witmore, Robin Valenza, and Jonathan Hope. The metadata has been further adapted by Deidre Stuffer ([stuffer@wisc.edu](mailto:stuffer@wisc.edu)) and Erin Winter ([ewinter3@wisc.edu](mailto:ewinter3@wisc.edu)) for the corpus release, under the supervision of Michael Gleicher ([gleicher@cs.wisc.edu](mailto:gleicher@cs.wisc.edu)). As part of the terms of the grant, we are happy to make the metadata freely available, though we ask that all use carry the above credit.

Although we have taken great care to check this data, there will inevitably be errors and information that is corrected by future research.

### General

- The metadata check involved checking the existing metadata (EEBO-TCP, ECCO-TCP) against information in the English Short Title Catalogue (<http://estc.bl.uk/>).
- In addition to the Helsinki Prototypical categories, Mattie Burkert generated a category ontology to address adapting the Helsinki categories to early modern and 18<sup>th</sup>-century British literature.

### Metadata Columns

- **TCP:** ID of the TCP file in which the text occurs
- **ESTC:** ID for the *English Short Title Catalogue*
- **text\_name:** name of plain text file, includes TCP ID
- **Title:** long title of the work
- **Title 30 Char or Fewer:** title given to work in 30 characters or fewer for easy identification purposes
- **Author:** recorded author of the text
- **Physical Description:** lists the size of the original text, whether it be a quarto (4 °), octavo (8 °), etc. Sizes for 1700+ content obtained first from *ECCO*. For pre-1700 material and broadsides, size was obtained from the *ESTC*.
- **Physical Description Notes:** indicates whether objects of 1-sheet had single-sided or double-sided printing. Entries are either 'single-sided' or 'double-sided.'
- **Derived Date:** year of the text standardized to the format YYYY.
- **Publication Place:** place of publication, exactly as it is in the text
- **Publisher:**
- **Verse, Prose, or Mixed:** The threshold for determining strictly verse and strictly prose: If a text is 80% or more prose, then it is labeled as prose. If a text is 80% or more verse, then it is labeled as verse.

- **Helsinki Prototypical Text Categories:** (<http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/textcategories.html>) Statutory (STA); Secular Instruction (IS); Religious Instruction (IR); Expository (EX); Nonimaginative Narration (NN); Imaginative Narration (IN). Applied to texts based on authorial intent and/or rhetorical situation.
- **Non-Helsinki Prototypical Categories:** Instructive (INST), Narrative (NARR), Expository (EX), Statutory (STA), Argumentative, Descriptive (DESC). This category was proposed to address when the Helsinki ontology was found wanting for categorization. This scheme combines both kinds of instruction, which is believed to fit the contours of print in the period better, where the fiction/nonfiction and religious/secular are amorphous distinctions. It also makes room to reintroduce argument and description, crucial categories that the Helsinki categories do not cover because they do not apply equally to Old, Middle, and Early Modern English. The Non-Helsinki Prototypical Categories were made by Mattie Burkert, Catherine DeRose, Katie Lanning, Mark Vareschi, and Elzbieta Beck.
- **Notes:** records issues noted by original metadata compilers, such as date discrepancies and where Helsinki Prototypical or Non-Helsinki Prototypical text categories are debatable
- **Status:** Status of document in TCP, either Free or Restricted

Deidre Stuffer

Last updated September 21, 2016