

7. Data Management Plan

The primary data that will be generated from this project will be the texts of the various anonymous and pseudonymous newspaper articles, as well as the analytic results of the authorship attribution tests. Retaining and publishing these will be the primary responsibility of the Papers of Abraham Lincoln. In particular, the newspaper articles will be published as part of the Illinois Digital Newspaper Collection (IDNC) at the University of Illinois at Urbana-Champaign. The articles that are assessed specifically as written by Abraham Lincoln will, in turn, be published as well as part of the Papers of Abraham Lincoln's Series II: Illinois Papers. The tools developed for analysis by the Evaluating Variations in Language Laboratory at Duquesne University will be published as per standard procedure through the lab website under open-source license.

In addition, we hope to re-use some of the data in other contexts. For instance, the PAN-CLEF conference, now in its third year, has been studying issues of authorship and authorship attribution. One decision taken at the most recent meeting (September, 2011, Amsterdam) was to focus on more real-world instances of disputed or unknown authorship to determine how computer-based analysis can interact with traditional scholarship and the relative advantages and disadvantages of both methods. The collection created by this project would be an excellent example of such an instance; republishing the datasets in such a framework would also help with the distribution and data maintenance issues.

Finally, of course, we expect this project to generate the usual collection of scholarly articles to help record the analytic results.