

"Scholars who work with manuscript materials encounter numerous difficulties beyond the obvious fact that handwriting can be difficult and ambiguous to read. In those instances where we are lucky enough to have large collections by a single author, the problem of dating the different stages of composition is often a significant issue. This problem is particularly pronounced in the case of early modern scientific and technological manuscripts, which were often separated or even dismembered and reassembled after the author's death, either by collectors, auction houses, or even by well-meaning librarians. Examples abound in the papers of Leonardo da Vinci, Galileo Galilei, Thomas Harriot, Robert Boyle, and Isaac Newton, to name only a few cases. A solution to the problems posed by dating large manuscript corpora would therefore be of benefit to researchers in many areas scholarship, not to mention librarians and museum workers. The present proposal aims to use Isaac Newton's very large manuscript Nachlass as a platform for exploring multiple techniques for dating manuscripts and for assessing them in combination with one another. These techniques include comparison of parallel passages generated automatically by a barrage of computational techniques (e.g. Latent Semantic Analysis and Topic Modeling), spectrometric analysis of inks and papers by Raman spectroscopy, comparison of watermarks by means of the latest imaging technologies, tracking of developments in orthography and handwriting, and the systematic study of citations, facilitated by the production of an authoritative electronic bibliography with hyperlinks to the citations in texts. In addition to applying these multiple techniques, the project will follow an integrative approach to assessing the data generated by them. In order to achieve this result, the project will use network graph analysis to produce visual clusters of manuscripts and passages whose respective chronological markers correlate with one another. Correlations generated by independent test-runs employing different techniques (e.g. ink-analysis and watermark-analysis) will then be compared to one another, using image graphs. In cases where a clean overlay of graphs does not result, team members will revisit the data and determine the cause of the poor 'fit.' Metrical techniques will also be developed for determining the degree of agreement that constitutes good versus poor fit. The result will be a model for determining relative chronology of the parts within large textual corpora that will be transportable to other projects."