New York's Dutch History:
Preparing a Discoverable Digital Resource from Primary Source Materials

presented by
Jeroen Dewulf,
Julie van den Hout,
& Mary Elings

Friday, November 17th, 2:00 PM - 4:00 PM
Academic Innovation Studio (127 Dwinelle)

In spring 2016, the German Department and the Bancroft Library partnered in a collaborative research grant through Digital Humanities at Berkeley to prepare a digital research collection from selected primary source materials in the Engel Sluiter Historical Documents Collection at The Bancroft Library. This unique research collection consists predominantly of copies and transcriptions of Spanish, Portuguese, Dutch, French and English primary source materials from archives in Europe, the United States, the Caribbean, and Latin America on the seventeenth-century Atlantic. These typed transcriptions of archival materials were previously inaccessible to most researchers because of difficulties in reading seventeenth century Dutch paleography. The project sought to design a web presentation for the "Colonial New Netherland" subset of documents, focused on the seventeenth-century Dutch colony of New Netherland, later, New York. The goal of the project was to digitize, extract, and clean the historic text, in order to present "research ready" text to enable natural language, machine-processing capabilities over these archival documents.

823 documents from the collection were digitized as TIFF files and then the digitized versions of the documents were run through Optical Character Recognition (OCR) to generate text files. The OCR text files were manually reconciled and corrected by way of the OCR Virtual Desktop supported by BRC's Analytic Environments on Demand (AEoD) service. The corrected texts were recombined into new PDF files, then run against web-based text analysis environment, "Voyant Tools," to explore the texts and determine if they were research ready. The results of the project were put into a website which presents the final research products, comprised of the corrected texts, presented as PDF files for use by researchers interested in doing text analysis over these archival documents. The text files can be used with other natural language processing tools, such as topic modeling, entity extraction, and keyword extraction, to explore and expand access to the documents. In addition to the project website presentation, the corrected texts are fully text searchable and published through Calisphere, a digital collection platform hosted by the California Digital Library.

Learn more and Register at dhfellows17.eventbrite.com

DIGITAL HUMANITIES
AT BERKELEY