
Beyond Cooccurrence: Network Visualization in the Civil War Governors of Kentucky Digital Documentary Edition

Patrick A. Lewis

patrick.lewis@ky.gov

Kentucky Historical Society, United States of America

Jeff Dycus

jeff.dycus@ky.gov

Kentucky Historical Society, United States of America

Anthony P. Curtis

tony.curtis@ky.gov

Kentucky Historical Society, United States of America

Whitney R. Smith

whitney.smith@ky.gov

Kentucky Historical Society, United States of America

Sara Carlsbad Brumfield

saracarl@gmail.com

Brumfield Labs, United States of America

Ben W. Brumfield

benwbrum@gmail.com

Brumfield Labs, United States of America

The Civil War Governors of Kentucky Digital Documentary Edition (CWG-K) is a freely-accessible online collection of historical documents associated with the chief executives of the state, 1860-1865. Yet CWG-K is about far more than the five governors. Uniquely positioned to view, reflect upon, and intervene in the lives of thousands of people experiencing the traumas of civil war, Kentucky's chief executives stood at the intersection of public and private life during an era that transformed the nation. Collectively, their papers provide a multi-faceted lens through which we can recover and understand more fully the lives of countless men, women, and children who have been heretofore both historically undocumented and archivally silenced.

After five years of active editorial work, CWG-K published the facsimiles and transcriptions of 10,000 documents [online](#) in 2016. Driven primarily by keyword searches and limited metadata faceting, *Early Access* is a digital evolution of the printed letterpress edition and its index. The content is interpretatively rich and suggestive and can be queried and sorted in a number of ways, but the experience is still ultimately linear. The ultimate goal of CWG-K, however, is to create a digital research environment within which a user can encounter the past multi-dimensionally through the documents and the powerful annotation network that links the documents together through the individuals, institutions, and places found in the texts.

CWG-K's true impact on scholarship, however, is through annotation. To the extent possible given the restrictions and biases of the historical record, CWG-K is identifying, researching, and linking together every person, place, and organization found in its documents. This web consisting of hundreds of thousands of networked nodes will dramatically expand the number of historical actors, show scholars new patterns and hidden relationships, and recognize the humanity and agency of historically marginalized people. The network of identified and annotated people, places, businesses, government agencies, and military units, will come as close as possible to a historical reconstruction of mid-nineteenth century society as it was lived and experienced in wartime Kentucky.

In this document-driven historical ecosystem, users can explore intuitively—moving seamlessly through seemingly disparate historical themes, events, and topics; breaking into the plane of social and geographic space to understand the deep patterns that underlay the issues raised in a text or set of texts; and moving forward and backwards through time to put this reconstructed historical world in motion and understand patterns of ebb and flow.

Phase II of the project (September 2016-September 2017) extends the edition into network visualization. Instead of reinventing the wheel for all the different functionality that an editorial annotation tool needs to have, we have taken the best existing open-source tools and integrated them to achieve our goals. TEI-XML files are exported from the transcription and control file tool DocTracker, published on the Omeka-powered Early Access website, and checked into a Github repository. Research assistants use Hypothes.is to identify entities within each document on the Early Access site. The custom-built MashBill tool queries Hypothes.is for those annotations, allowing research assistants to identify references to entities and

build relationships between initially cooccurring entities and document biographical research from external sources. MashBill then uses the annotations and identifications to update the TEI documents automatically with the appropriate persName/orgName/placeName references and re-publishing them to Github.

In the open source tool MashBill, CWG-K makes an important contribution to network analysis in digital humanities. With a few exceptions network analysis is still dominated by network construction based on cooccurrence of entities within documents. These entities are most frequently created by automated named entity recognition performed on plaintext. CWG-K researcher's close reading of documents to extract entities allows for much accurate identification of entities. The use of Hypothes.is makes such human-powered entity recognition far more scalable than traditional manual tagging of TEI-XML. Rather than using pure cooccurrence MashBill allows relationships to be defined by researchers consulting resources outside the documents of the edition. These relationships are then visualized with a D3.js visualization which is deeply linked to both the documents themselves as well as the articles written during the course of entity and relationship research on the people, places, and things. The open source MashBill tool may be reused for any TEI/Omeka project to reduce the effort and improve quality of entity identification.

Since the Civil War Governors of Kentucky Digital Documentary Edition is a project of the Kentucky Historical Society, it has always focused on public access. The development of MashBill and integration of the network visualization produced with that tool into the early access website will enable the public to discover the lives and stories of everyday people who interacted with the offices of the governors. Our synthesis of approaches and technologies provides an example other projects can benefit from, showing how to leverage open source tools and standards to efficiently identify and build network visualization in public digital editions.