



Presentation Details
Title: Citation Preservation as a Tool to Open Paywalled Sources to Computational Analysis for Novice Digital Humanities Practitioners
Speaker: Prof. Kalani CRAIG (Indiana University Bloomington)
<p>Abstract:</p> <p>Digital humanists first encountering text mining often assume their project will benefit from automated language processing on open-source freely accessible data sets, clean and ready to be adapted from one project for use in another. The reality of the data landscape for digital humanists is much more complicated. Many digitized, transcribed sources available for use in text mining are in English, institutionally owned and often paywalled. Additionally, archives often place limitations on the transcription results of their archived documents. These copyrighted digitized texts are still valuable targets for text mining, HGIS and network-theoretical approaches, but it can be difficult to provide results that meet open-data best practices within the constraints of the original copyrights. Finally, the theory and practice of computational linguistics is often well outside the standard training for humanists embarking on a digital project.</p> <p>This paper uses case studies in two different languages — memory building in the medieval Latin of saints' lives and an exploration of the language of spying in the Cuban War of Independence — to demonstrate the need for a practical documented process that bridges the gap between novice digital humanists working outside of English-language environments and more advanced computational linguistics. This process yokes best practices in data curation to best practices in computational linguistics in order to manage the preservation of citation data, which is often lost or obscured as we work with topic modeling and natural-language processing. The resulting citation practices allow digital humanists to put their work into dialogue with humanists working in more traditional spaces. A set of online documentation and resources for replicating the process will accompany the paper.</p>
<p>Biography</p> <p>Kalani Craig is Clinical Assistant Professor in the Department of History at Indiana University Bloomington. Her research and teaching specialties are at the intersection of digital methodologies and</p>

medieval history. She uses text mining, spatial history and network analysis to explore a variety of histories, including her own focus on conflict in medieval European history. Her in-classroom work makes similar use of these computational approaches. Prior to returning to the scholarly world, Dr. Craig spent ten years managing Web sites and the technical and creative people responsible for creating those Web sites.

She and co-author Arlene Diaz have recently signed a contract for a monograph entitled “Scientific Knowledge in the Spanish-Cuban-American War: Spies, Detectives, and Digital History” with Indiana University Press. Dr. Craig's other publications include articles on the use of computational analysis in humanities classrooms, the most recent of which was “Analog Tools in Digital History Classrooms: An Activity-Theory Case Study of Learning Opportunities in Digital Humanities” in the *International Journal for the Scholarship of Teaching and Learning*. She is a Mosaic Faculty Fellow at Indiana University in 2015-17 and since 2015 has served as a committee member in the American Historical Association's Digital History Working Group, which advocates for the development and evaluation of digital scholarship in the history discipline across the United States.