

In the digital era, the development of computational tools and mass-scale digitization raise new opportunities and challenges in dealing with cultural heritage data and archives. Millions of pages of the VOC (“De Vereenigde Oostindische Compagnie”, known as the Dutch East India Company in English) archives preserved in the Netherlands and Asia have always been the sources for the study of the history of the Company itself. However, the VOC archives as perhaps the largest corpus about the Orient produced by a single organization in the seventeenth and eighteenth centuries record not only information about the Company’s mercantile, maritime, diplomatic, administrative, and colonial activities in Asia but also about local societies and people with which the Company encountered. In recent debates over the (further) explorations of the VOC archives in the digitization context, scholars argue for the possibilities of re-reading and re-evaluating the VOC archives from interdisciplinary perspectives, and of rethinking and rewriting the VOC history from non-Dutch-centred, cultural, and post-colonial approaches. Such debates not only pose the question of what computational method and digital tool shall be (further) developed to fulfil the need to analyze the cultural heritage data efficiently but also motivate us to examine the impact of digitization, computational methods, and digital tools on the practice of historical research.

The digitization of the VOC archives has lately made significant progress as the digital infrastructure project GLOBALISE has generated transcriptions of the c. 5 million handwritten pages of the VOC archives and in October 2023 launched a Transcriptions Viewer for users to digitally search information and explore topics such as enslavement and colonialism in the archives. In this paper, I investigate how computational methods such as natural language processing and knowledge representation can be used to (re)discover the under-studied topics in the VOC archives and to reveal the hidden voices of the indigenous people in Asia in the VOC context. My hypothesis is that automatically retrieving information about an entity (e.g., one ethnic group in Asia) or an event (e.g., the 1740 Batavia massacre) that is collected in different places from different perspectives over time and then integrating the information into a knowledge graph will equally show the possible diverse dimensions of an entity or an event in the VOC context. By showing these dimensions and voices, my research can aid scholars in exploiting the potential that the digitized VOC archives have for cultural analysis, regional and global history, as well as for decolonization. My research could also serve as an example of evaluating to what extent digitization and computational methods can bring opportunities to the historiographical practices of the VOC archives.