Title: The Ethics of Quantifying Holocaust Archives

Author: Alexis Lerner

Abstract: Archival materials about the Holocaust are, at once, substantial and understudied. One reason for this is the tremendous breadth of each collection. For example: the Arolsen Archives, formerly the International Tracing Service Archives, hold 30 million original documents about Nazi crimes and displaced persons. The files—ghetto inhabitant lists, deportation records, photographs of personal items collected in concentration camps—could stretch for an alleged 16 miles (Shapiro, 2007). Likewise, there are approximately 55,000 videotaped testimonies of Jewish Holocaust survivors in the USC Shoah Foundation's Visual History Archive. Yale University's Fortunoff Archive has 4,400 videotaped Holocaust testimonies, and Indiana University's AHEYM archive has another 400.

As scholars from the social sciences and mathematics reinterpret these archives, they apply quantitative methods in order to systematize scholarly analysis of this so-called 'big data'. For example, Knowles employed historical GIS to map concentration camps, Lerner used text analysis to code the confluence of emotional responses and national institutions in testimonies, and Lee designed a machine learning algorithm to digitize multilingual death certificates. What are the ethical implications and limitations of these methodological decisions? For example, is there a reductionist tendency in quantifying individual or collective traumas for the purpose of scholarly research? If so, how can researchers best communicate the uniqueness of each observation—every individual artifact or testimony in their dataset—to their readers? Quantitative methods can revolutionize our understanding of the Holocaust, but threaten to sacrifice personal narratives in exchange for predictive capabilities; as such, clear ethical considerations regarding the quantification of Holocaust archives are crucial.