

# Creating Primary Digital Archives Via Web Archiving Techniques

ALEXANDER DURYEE

**Abstract:** The digital age has created an abundance of primary source collections in nearly every field. While few of these are in formal controlled archives, they represent an extraordinary research resource. By applying cutting-edge digital humanities tools to these corpora, researchers can explore previously invisible topics and relationships. However, despite the promise of such collections, very few are readily available for research - issues in formatting, formats, and location serve as powerful barriers to use.

In this session, I will discuss my process and results of acquiring and preparing the Archives and Archivists mailing list (1993-2013) for use in algorithmic text analysis. I will explore issues in harvesting data "from the wild" and methods used to prepare it for analysis. By exploring the process in adapting digital materials to algorithm-friendly formats, I will discuss its implications for digital archives, notably with regard to acquisition and in opening collections to the future of research. Additionally, I will apply text analysis methodologies to derive otherwise invisible conclusions about the listserv and profession, as examples of the research potential of uncontrolled archival corpora.

## **About the author:**

*Alex Duryee* is a digital preservationist and metadata specialist with AudioVisual Preservation Solutions. His experience and interests include web archiving, optical and magnetic storage migration, and data preservation/management. Holding an MLIS from Rutgers University, his background includes archival tool development, data access, preservation infrastructure analysis, and metadata management. His past work includes digital conservation with Rhizome, where he preserved and revived net- and disc-based artworks.