ArchivesZ: Visualizing Archival Collections
Tackling the Challenges of Data Aggregation


**RESEARCH GOALS**
Extract and aggregate collection level data from EAD encoded finding aids. Generate cross-collection and cross-repository visualizations.

**BUILDING BLOCKS**
- **Subjects**
- **Linear Feet**
- **Years**

What? How much? When?
Data extracted from XML encoded descriptions of archival collections

**TARGET AUDIENCES**
- Archivists and manuscript curators
- Researchers, historians and humanities scholars
- Students

**METODOLOGY**
- Encoded Archival Description (EAD): international standard for XML encoding descriptions of collections
- EAD XML files acquired from many repositories
- Extract, normalize and transfer data to database

Decompose SUBJECTS to TAGS
Remove redundant and non-descriptive terms

Agricultural colleges – Maryland – History – Sources
Tobacco – Maryland – History – Sources

**CHALLENGES**
- Flexibility of EAD encoding guidelines
- Diversity of descriptive practices and XML encoding standards across institutions
- Wildly diverse subject terms, both within and among repositories
- Collection level subjects associated with all records
- Conversion of all sizes to linear feet

**NEXT STEPS**
- Collection of additional finding aids
- Refinement of data extraction logic
- Support update of finding aids
- Programmatic removal of stop word tags
- Institution level configuration files
- Refinement and completion of Version 2 visualizations

**VISUALIZATIONS OF AGGREGATED DATA**

**ArchivesZ Project Credits:**
Version 2 Team: Jeanne Kramer-Smyth, Richard Bovell
Version 2 Project Funded by NEH Digital Humanities Startup-Grant
NEH Project Director: Dr. Jennifer Golbeck
Version 1 Team: Jeanne Kramer-Smyth, Tim Anglade, Morimichi Nishigaki

- Heat Map Grid: Tags vs Decades
- Dual Histogram: Tags vs Size
- Stacked Time: Tags vs Decades
- Histogram: Decades vs Size

**VISUALIZATIONS OF AGGREGATED DATA**

Civil Rights
Education
Art
Agriculture

MARYLAND'S iSCHOOL in the INFORMATION CAPITAL of the WORLD