# Basic Finding Aid Clean-Up Using OpenRefine

Presented by: Bailey Hoffner, Metadata and Collections Management Archivist

UNIVERSITY LIBRARIES
*The UNIVERSITY of OKLAHOMA*

*What will you do at the intellectual crossroads of the University of Oklahoma?*

# What is OpenRefine?

An open-source, browser-based tool for cleaning up messy data

https://openrefine.org

*What will you do at the intellectual crossroads of the University of Oklahoma?*
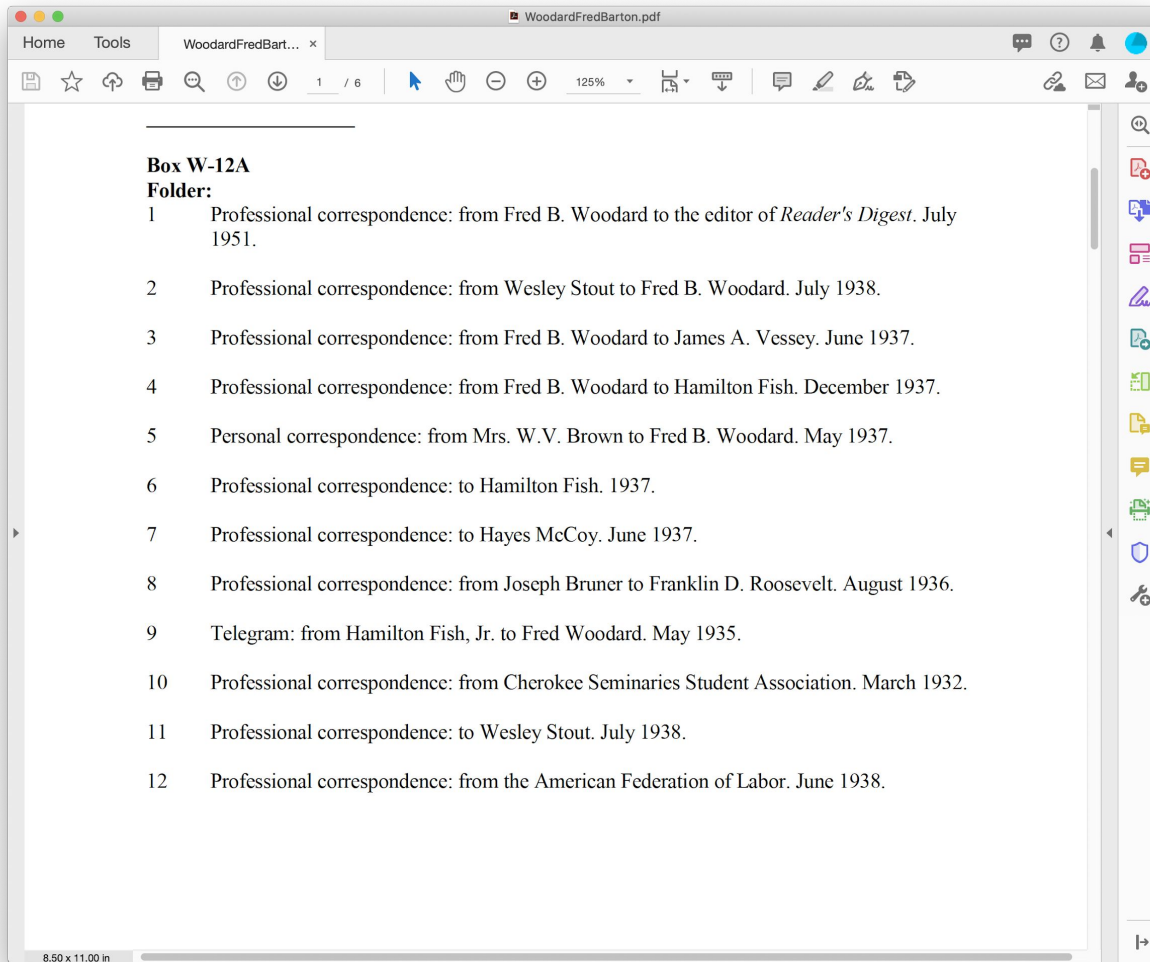
# Why OpenRefine for Archivists?

- low-touch approach to transforming legacy finding aids

- batch fix human error

- encourages understanding of finding aids as iterative documents

- structured data = more options in the future

*What will you do at the intellectual crossroads of the University of Oklahoma?*

# Goal: Structured Data

# PDF → Spreadsheet

# OpenRefine: Initial Formatting Options

# OpenRefine: Trim Whitespace

# OpenRefine: Trim Whitespace

# OpenRefine: Split Columns

# OpenRefine: Split Columns

# OpenRefine: Split Columns

# OpenRefine: Cluster & Edit

# OpenRefine: Cluster & Edit

# OpenRefine: Regular Expressions

# OpenRefine: Regular Expressions

# OpenRefine: Join Columns

# OpenRefine: Join Columns

# OpenRefine: Join Columns

# OpenRefine: JSON Export & Reuse

Live Walk Through!

*What will you do at the intellectual crossroads of the University of Oklahoma?*

# OpenRefine: JSON Export & Reuse

# Thank You

Bailey Hoffner

Metadata and Collections

Management Archivist

baileys@ou.edu

*What will you do at the intellectual crossroads of the University of Oklahoma?*