



# Maximizing Productivity: Expediting Metadata Creation and Cleanup for Electronic Theses and Dissertations at Oklahoma State University



Madison Chartier, Digital Resources & Discovery Services, Oklahoma State University

## Introduction

OSU's ETD metadata creation workflow depended on manually intensive procedures. With only one metadata librarian, this workflow needed revision to automate batch processes. Revision included reassessment of the tools used, which impacted the steps required, for an expedited, automated processing workflow.

## Materials and methods

Students submit ETDs through UMI ProQuest. Preliminary metadata is created and submitted as an XML file.

The original workflow depended on:

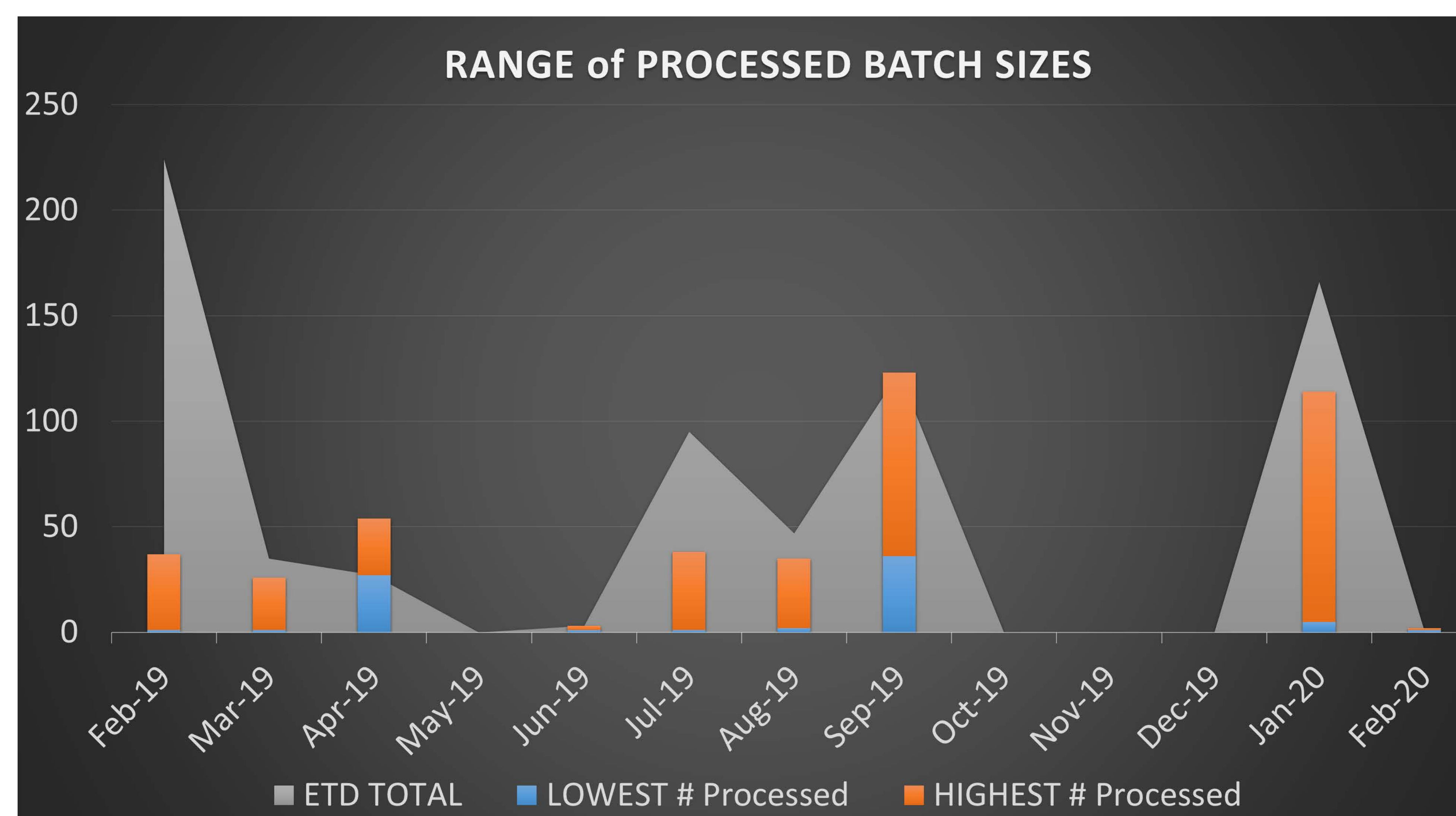
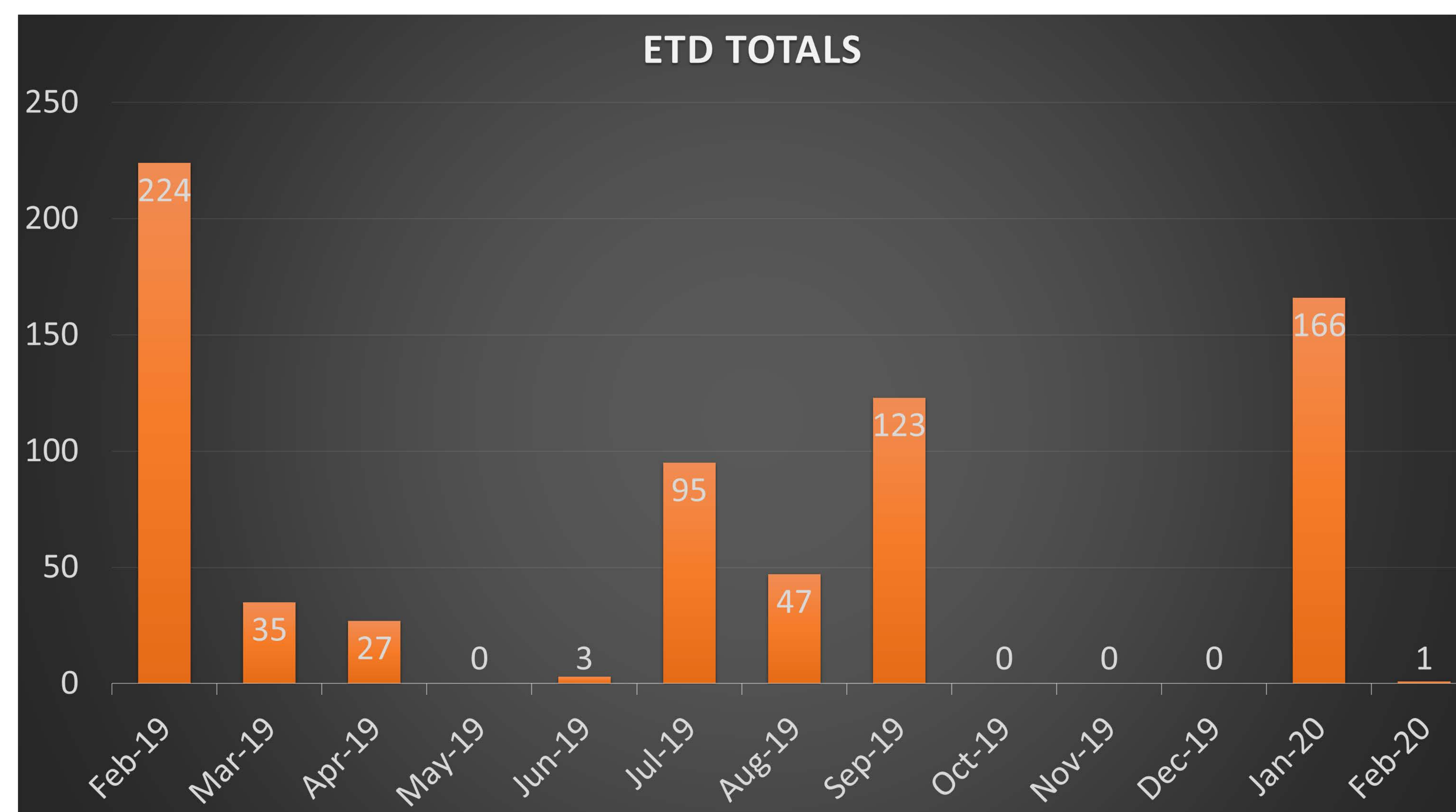
- Obsolete tools
- Copy and paste methods to facilitate:
  - XSLT functions on XML files
  - Compilation of all metadata in one file

The revised workflow depends on:

- XSLTs
  - #1: Extract and format values
  - #2: Compile all metadata in one XML
- OpenRefine
  - Metadata cleaning
  - XML → CSV file conversion

## Results

The revised workflow, introduced in February 2019, automated metadata creation and cleanup for **664** ETDs over the past year:



## Conclusions

Metadata creation and cleanup for newly processed ETDs shrank from complex, manually-intensive procedures to a seamless automated workflow. ETD batches *exceeding 100 entries* have metadata created, cleaned, and compiled for upload in *under 10 minutes*.

The combination of **open-source, actively-maintained, widely-used** technologies guarantees longevity of workflow.

Human endeavor is reduced to:

- verifying accuracy and completeness of metadata;
- documenting embargo dates for future upload; and
- file archiving.

## Acknowledgments



Oxygen XML Forum, "Merge Multiple Files into One XML"

## Further information



madison.chartier@okstate.edu