Capitalizing on Automated Workflow

Expediting Metadata Creation for Electronic Theses and Dissertations at Oklahoma State University

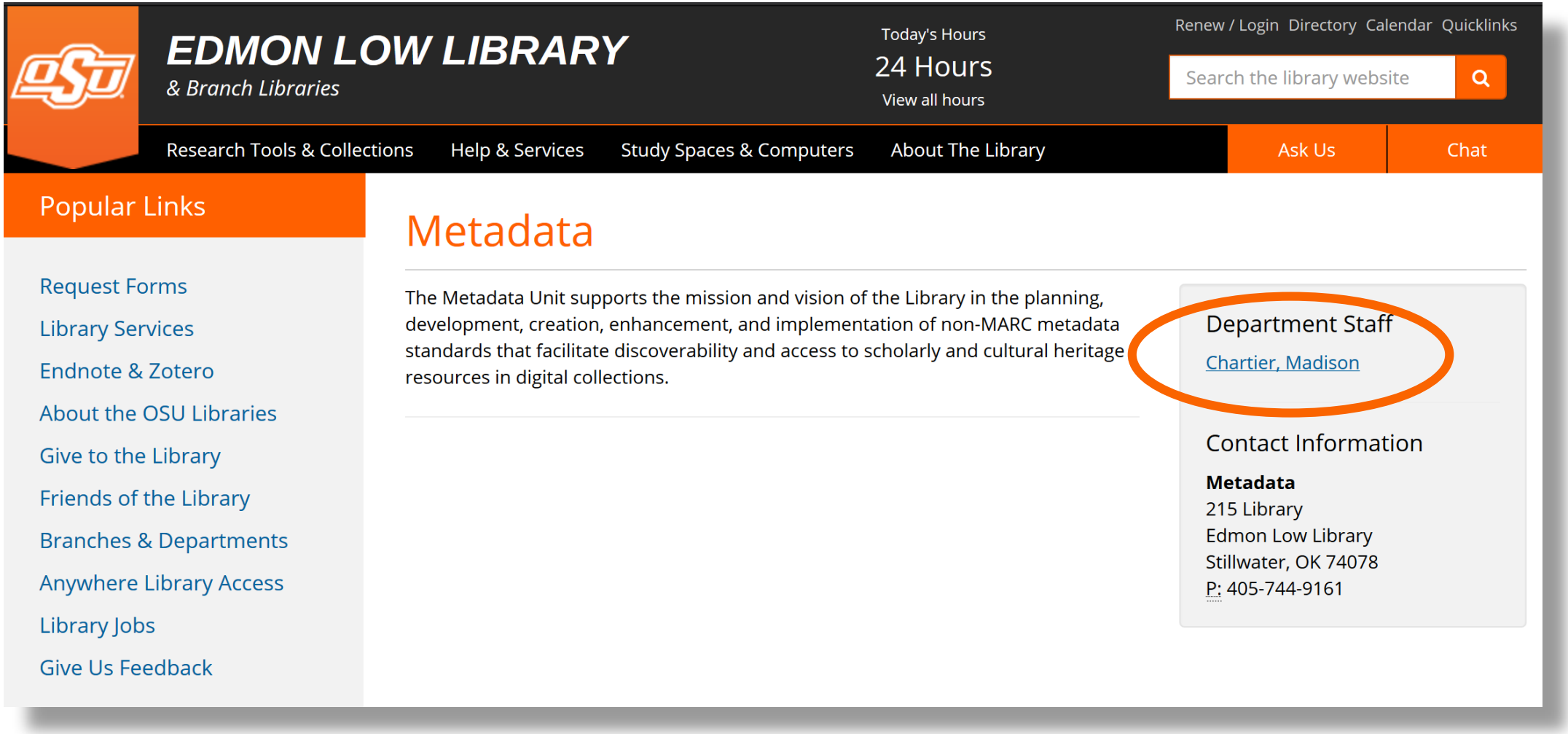
MADISON CHARTIER
Metadata Librarian, Oklahoma State University
Oklahoma Library Association
Technical Services Round Table Workshop
30 October 2020

Outline

Context

OSU's Metadata Team
OSU's ETD Collection
UMI ProQuest

OSU's Metadata Team



Sparse, lean, & mean.

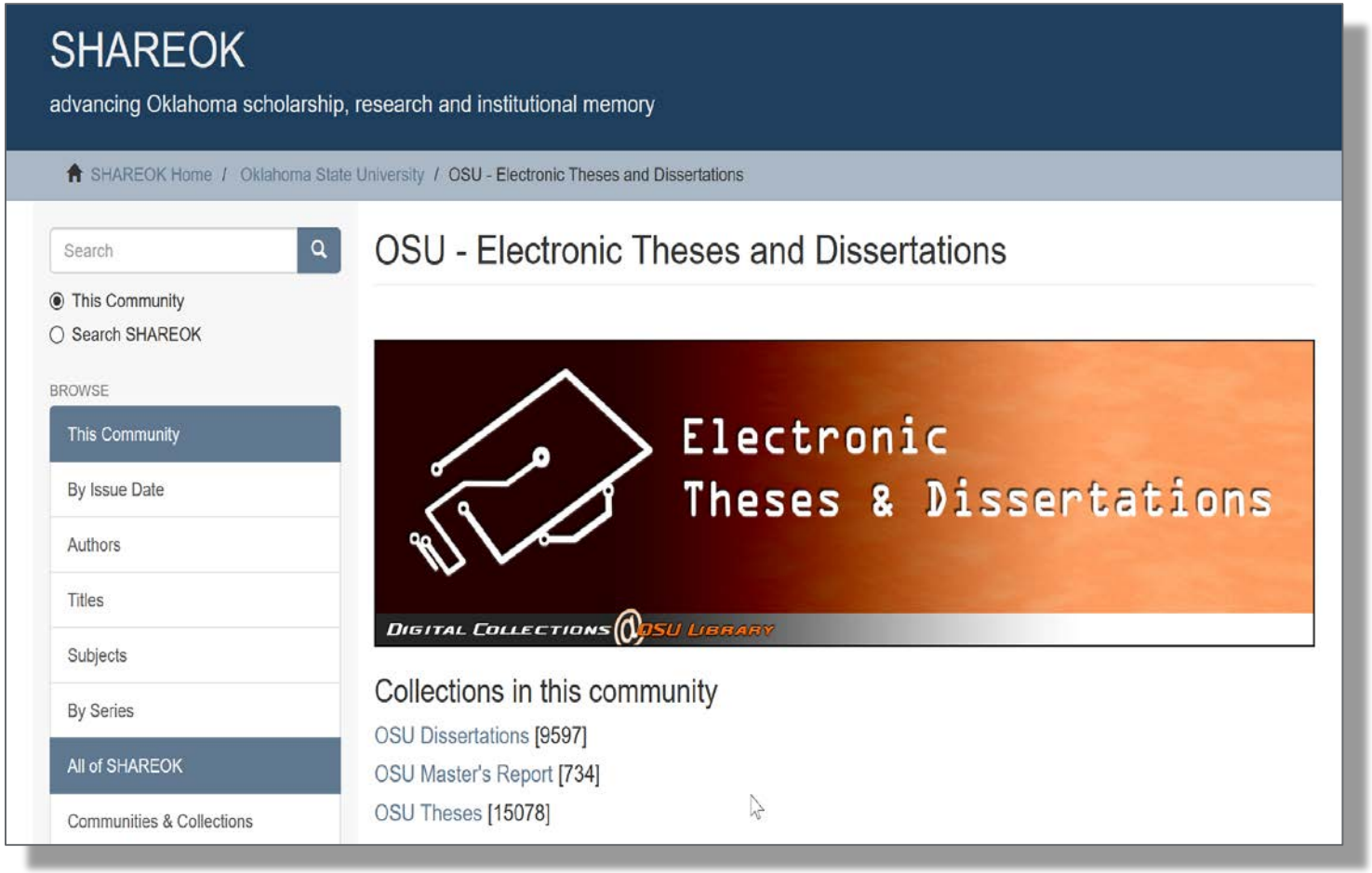
OSU's ETD Collection

OSU's largest online collection

- **25,400+** items
- 2019:
  - **554** new entries
  - **36** separate uploads
  - Max. of **96**; min. of **1**

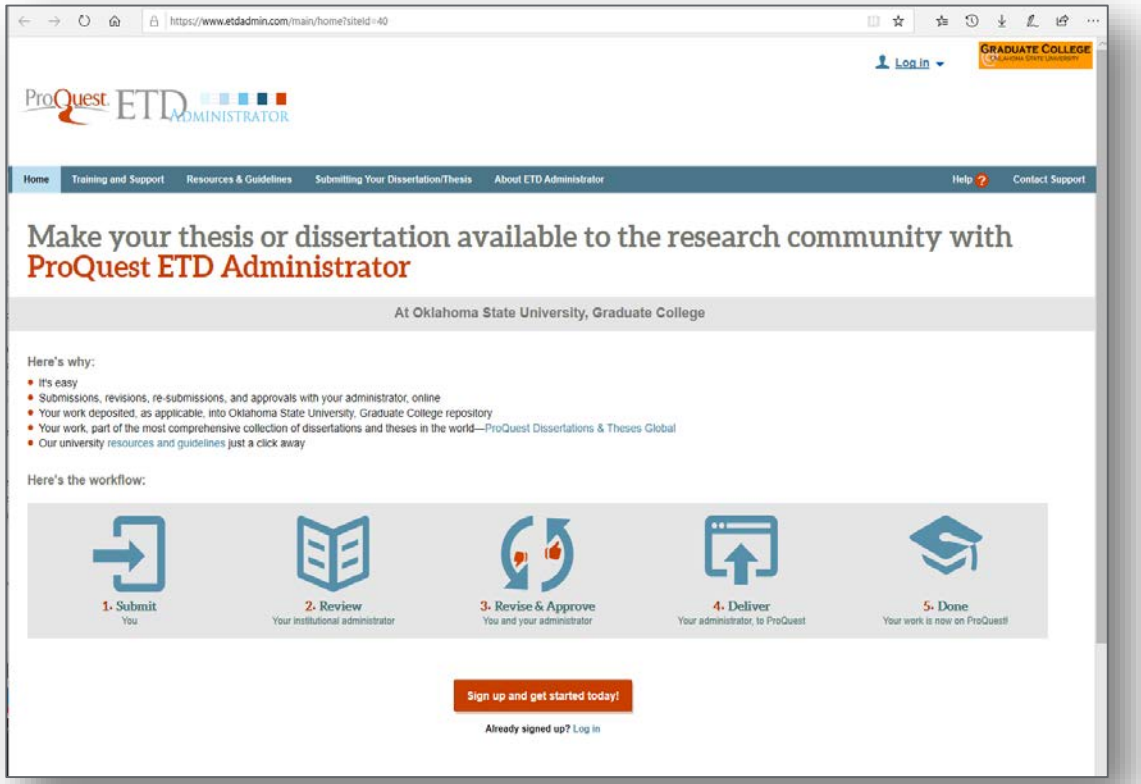
High-priority collection

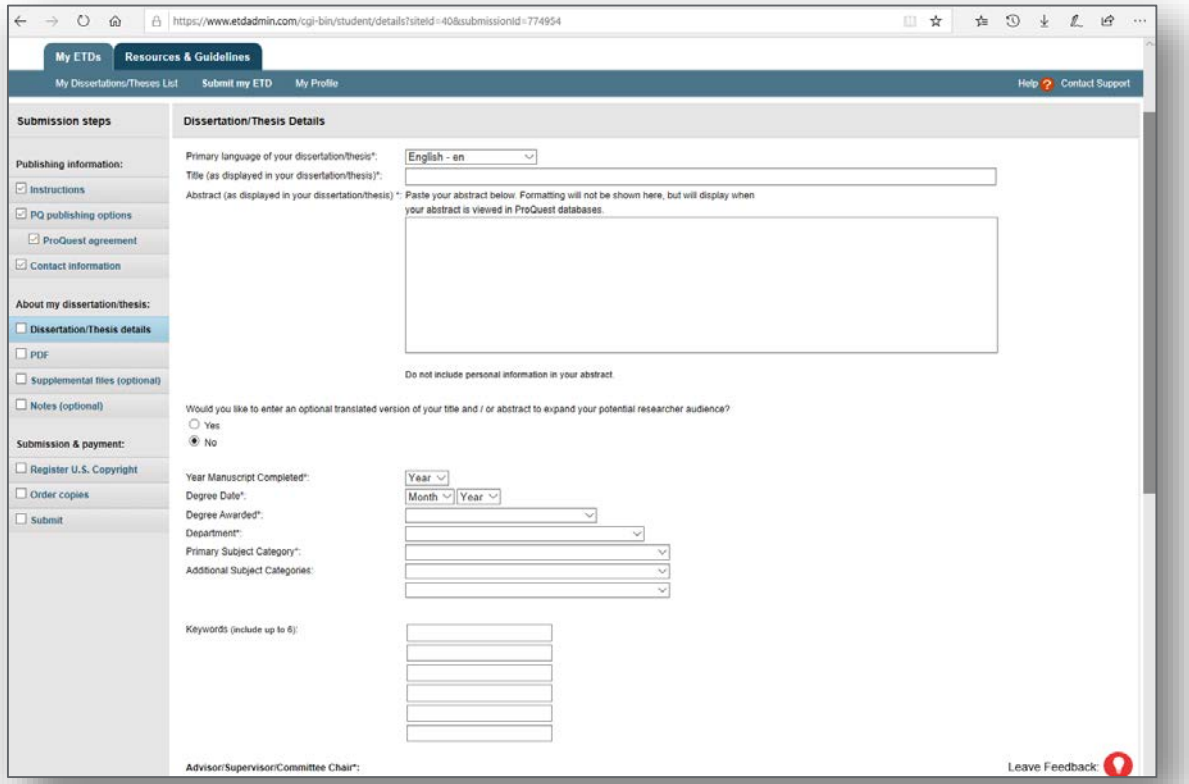
- Student-faculty work is the pride of OSU!
- We process & publish ASAP upon ingest!



A tall order for a single metadata librarian…

UMI ProQuest





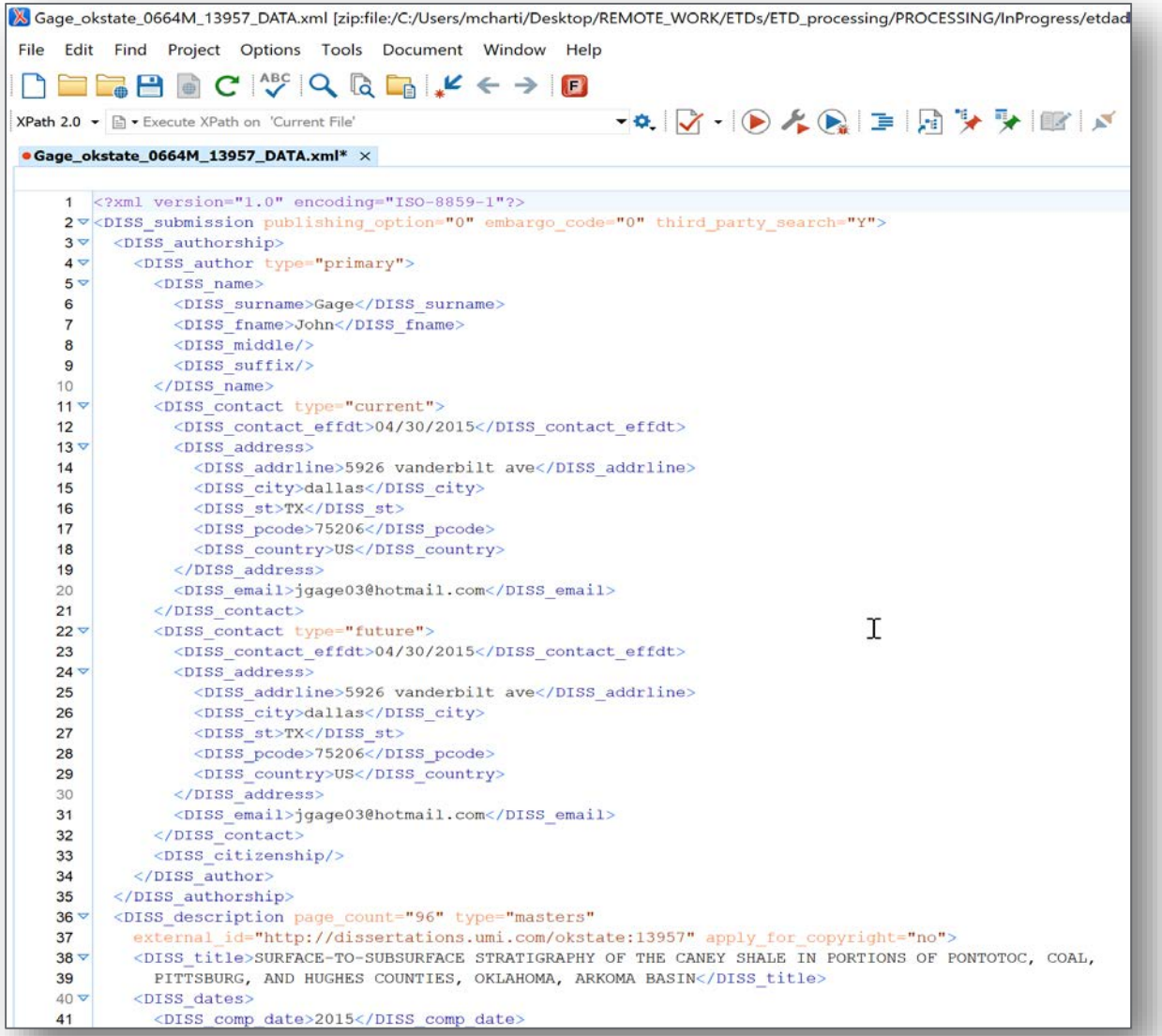
Images provided by Dr. Jean Van Delinder, Senior Associate Dean of the Graduate College.

Consistent, uniform metadata → **AUTOMATED BATCH PROCESSES!**

Consistent XMLs

Consistency enables everything!

- Same fields/tags
- Same value formats
- Individual values for each ETD
- Every item's metadata is set up the same way
- Permits batch processes & automation



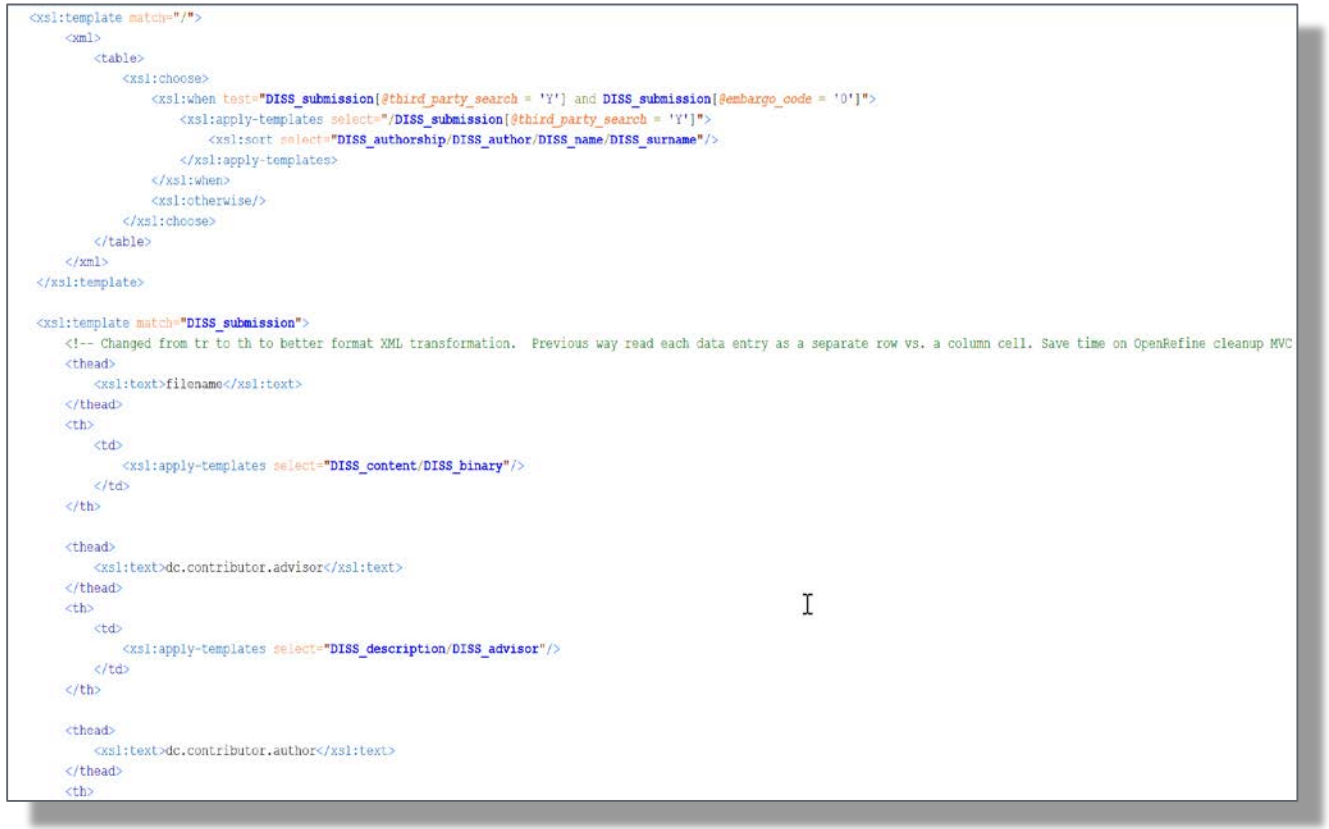
Consistency → Batch Processing → Automation

Workflow Tools

eXtensible Stylesheet Language Transformations
OpenRefine

XSLT #1: Metadata Extraction & Prep

1) Pulls and formats metadata from XML files
2) Assigns Qualified Dublin Core field labels
3) Sorts out embargoed and restricted entries
4) Generates 2nd XML file with prepped metadata



Preps metadata to meet standards established in the collection's Metadata Application Profile.

XSLT #2: Metadata File Merger

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:xs="http://www.w3.org/2001/XMLSchema" exclude-result-prefixes="xs" version="2.0" xmlns:fn="http://www.w3.org/2005/xpath-functions">

    <!-- Below is a recommended insert from Oxygen XML Editor online: https://www.oxygenxml.com/forum/topic11375.html.
        This is meant to help with merging the multiple XML files into one.

        This XSLT serves as STEP 2 in the ETD workflow.  The following code should take all of the XML files in a batch (make sure they're separated

    <xsl:output method="xml"/>
    <xsl:template match="/">
        <xsl:copy>
            <xsl:apply-templates mode="rootcopy"/>
        </xsl:copy>
    </xsl:template>

<!-- PROBLEM: in this section.  Makes multiple copies. Copies all files 30 times.  ??? -->
    <xsl:template match="node()" mode="rootcopy">
        <xsl:copy>
            <xsl:variable name="folderURI" select="resolve-uri('.', base-uri())"/>
            <xsl:for-each
                select="collection(concat($folderURI, '?select=*.xml;recurse=yes'))/*/node()">
                <xsl:apply-templates mode="copy" select="."/>
            </xsl:for-each>
        </xsl:copy>
    </xsl:template>

    <!-- Deep copy template -->
    <xsl:template match="node() | @*" mode="copy">
        <xsl:copy>
            <xsl:apply-templates mode="copy" select="@*"/>
            <xsl:apply-templates mode="copy"/>
        </xsl:copy>
    </xsl:template>

    <!-- Handle default matching -->
    <xsl:template match="*"/>

</xsl:stylesheet>
```

Merges individually prepped files into one mass file.

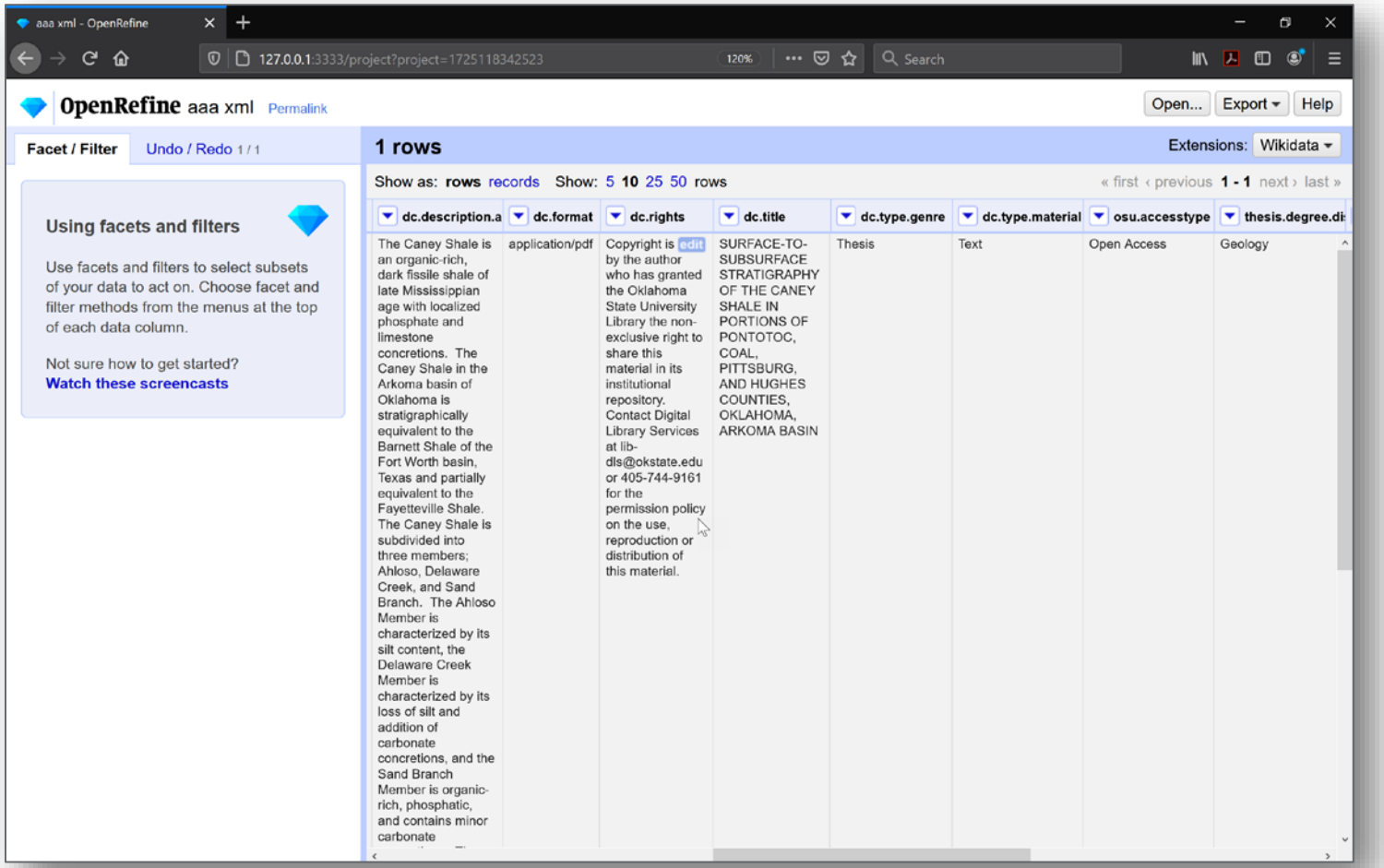
OpenRefine

Open-source tool in popular use at OSU Library

- Cleans messy data!
- Use for metadata cleaning

Enables our file conversion!

- Can upload compiled XML file…
- …clean metadata…
- …and export as a comma-separated values (CSV) file!



All that remains = accuracy checks against PDF files!

Live Demo

OSU's ETD Processing Workflow

Summary

Madison Chartier

Digital Resources & Discovery Services

O | 405.744.9536
E | madison.chartier@okstate.edu
215A Edmon Low Library

https://info.library.okstate.edu/c.php?g=916329

THANK YOU!