

# Legacy ETDs with Legacy Problems

## Changing ETD Workflows to Improve Discoverability

### Abstract

Atkins Library receives ETD metadata records from ProQuest, delivered in Proquest's XML standard. This metadata includes student-supplied keywords and controlled terms from ProQuest's own subject vocabulary—but no terms from the FAST subject vocabulary used in Atkins' institutional repository, Niner Commons. This omission prevented subject faceting and collocation both in Niner Commons and in the library's catalog. Student keywords were also in all caps, which is inconsistent with widely accepted standards for metadata display. This remediation project applied FAST terms to ETD records, normalized capitalization, and eliminated whitespace and formatting errors to improve discoverability and the user experience.

### Tools

- OpenRefine
- FAST reconciliation [code](#)
- Google Sheets
- [XSLTs](#)
- Islandora repository software

### Steps

- 1) Reconcile.** Atkins staff extracted ProQuest subject terms from its legacy ETD records and reconciled them in OpenRefine against the FAST subject vocabulary. FAST matches were evaluated for accuracy as well as for adherence to new local metadata standards for diversity, equity, inclusion, and accessibility. Substitutions from alternative vocabularies such as Homosaurus were made where appropriate.
- 2) Create.** FAST vocabulary terms and their associated URIs were exported from OpenRefine and incorporated into an XSLT that applied the reconciled terms to legacy MODS repository records for ETDs based on matches with ProQuest subject terms already in the records.
- 3) Remediate.** Atkins staff extracted all legacy ETD records from the repository, remediated the records with the XSLTs, and then ingested the ETD records into the repository.
- 4) Revise.** Atkins staff revised the XSLT used for transforming incoming ETD records so that the sme FAST subject terms and style standards would be applied to newly arriving ETD records from ProQuest.

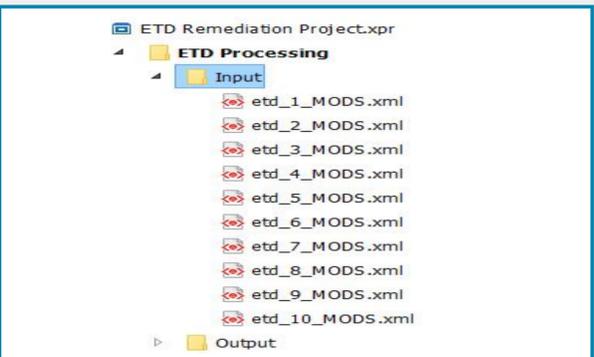
### Workflow

ProQuest Keyword	Terms to Reconcile	URI
ASIAN AMERICAN STUDIES	Ethnicity--Study and teaching	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
ASIAN AMERICAN STUDIES	Asian Americans	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
BALTIC STUDIES	Baltic States	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
BALTIC STUDIES	Area studies	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
CANADIAN STUDIES	Canada	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
CANADIAN STUDIES	Area studies	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
EUROPEAN STUDIES	Europe	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
EUROPEAN STUDIES	Area studies	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
FRENCH CANADIAN CULTURE	Ethnicity--Study and teaching	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
FRENCH CANADIAN CULTURE	French-Canadians	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
HOLOCAUST STUDIES	Jewish Holocaust (1939-1945)	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
NORTH AFRICAN STUDIES	North Africa	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>
NORTH AFRICAN STUDIES	Area studies	<a href="http://id.worldcat.org/fast/1009532">http://id.worldcat.org/fast/1009532</a>

1. Extract ETD records from Niner Commons
2. Perform reconciliation against FAST in OpenRefine

```
<xsl:if test=".='Naval engineering'">
  <mods:subject authority="fast" xmlns="http://www.loc.gov/mods/v3"
    valueURI="http://id.worldcat.org/fast/1009532">
    <mods:topic>Marine engineering</mods:topic>
  </mods:subject>
</xsl:if>
<xsl:if test=".='Public policy'">
  <mods:subject authority="fast" xmlns="http://www.loc.gov/mods/v3"
    valueURI="http://id.worldcat.org/fast/1353198">
    <mods:topic>Government policy</mods:topic>
  </mods:subject>
</xsl:if>
```

3. Use OpenRefine template to move keywords and URIs into XSL "If" elements
4. Build out XSLT to replace terms



4. Apply XSLTs using Oxygen project function
5. Replace Niner Commons ETD records with revised files

### Results of transformation in sample ETD record

Before

Genre	<a href="#">doctoral dissertations</a>
Degree	<a href="#">Ph.D.</a>
Keywords	<a href="#">COMMUNITY BENEFIT</a> <a href="#">COMMUNITY HEALTH</a> <a href="#">COMMUNITY HEALTH NEEDS ASSESSMENTS</a> <a href="#">POPULATION HEALTH</a> <a href="#">TAX-EXEMPT HOSPITALS</a>



After

Genre	<a href="#">doctoral dissertations</a>
Subjects--Topics	<a href="#">Public health</a> <a href="#">Government policy</a> <a href="#">Health services administration</a>
Degree	<a href="#">Ph.D.</a>
Keywords	<a href="#">Community Benefit</a> <a href="#">Community Health</a> <a href="#">Community Health Needs Assessments</a> <a href="#">Population Health</a> <a href="#">Tax-Exempt Hospitals</a>

### Challenges

- Duplicate terms:** the XSLT to crosswalk terms to FAST introduced duplicates for a few terms, necessitating a second XSLT to eliminate duplicates.
- Implementing changes:** our repository, which runs on Islandora, was not ingesting the remediated metadata as the repository was overdue for a system upgrade. We worked extensively with IT to rectify. There were also significant processing delays when applying the XSLTs to the records.
- Offensive and outdated terms:** reviewing subject terms uncovered offensive and outdated terminology, spurring us to establish new cataloging guidelines.

### Looking ahead

- Additional remediation:** Projects include reconciling ProQuest-provided subject areas to UNC Charlotte departments to enable department browse and collocation.
- Address student-supplied keywords,** which are uncontrolled and thus variable quality. Some are misspelled, others are overly specific and apply to single ETDs only. May want to create parameters and/or perform some basic clustering and cleaning.