

# Migrating ETDs from Dublin Core to MODS: *Automated processes for metadata enhancement*

ALCTS Metadata Interest Group  
Virtual Pre-Conference  
June 7, 2016



# Migrating ETDs from Dublin Core to MODS



**Annie Glerum**

Head of Complex Cataloging  
Florida State University

[aglerum@fsu.edu](mailto:aglerum@fsu.edu)

<https://github.com/aglerum>



**Dominique Bortmas**

Metadata Librarian  
University of South Florida

[dbortmas@usf.edu](mailto:dbortmas@usf.edu)

<https://github.com/dbortmas12>



## Survey

How familiar are you with ETD metadata  
(Choose one)

- ◆ I work with ETD metadata often
- ◆ I work with ETD metadata occasionally
- ◆ I work with ETD metadata rarely
- ◆ I know about ETD metadata but don't work with it
- ◆ I don't know anything about ETD metadata

## ETD Migration

- ◆ ETDs at FSU
- ◆ Thin vs. Thick Metadata
- ◆ Harvesting and Refining PDF Data
- ◆ Modular XSLT Transformations
- ◆ Multiple XML Sources

## Adaptations

- ◆ Institutional Repository metadata to MARC21
- ◆ Generating NACO List



# ETD Migration



## ETD Workflows at FSU

**Pre-2012:** Student submits to Graduate School

- ◆ Metadata was entered manually

**2012-2015:** Student submits to ProQuest

- ◆ ProQuest metadata to Bepress

**2015:** Bepress migrated to Islandora

- ◆ Bepress metadata to MODS

**2016:** DigiNole Research Repository

- ◆ ProQuest metadata to MODS



# Thin vs. Thick Metadata



# Migrating ETDs from Dublin Core to MODS

## Thin Metadata

Title  
Publication Date  
Author Name  
Institution Names (University, Dept.)  
Keywords  
Abstract  
Advisory Committee Note  
Degree Name

## *Thick Metadata*

Title  
Author Name  
Committee Member Names  
Institution Names (Univ., College, Dept.)  
Publication Date  
Language  
Physical Description  
Abstract  
Submitted Note  
Degree Awarded Note  
Date of Defense Note  
Keywords  
Bibliography Note  
Advisory Committee Note  
Controlled Subjects  
Degree Name



# Migrating ETDs from Dublin Core to MODS

## Core Description

Text string

`<title/>`

## ***Full Description***

Parsed values

```
<mods:titleInfo>  
  <mods:title/>  
  <mods:subTitle/>  
  <mods:nonSort/>  
</mods:titleInfo>
```



# Migrating ETDs from Dublin Core to MODS

## Core Description

Text string

```
<fields>  
  <field name=  
    "degree_name">  
    <value/>  
  </field>  
</fields>
```

## ***Full Description***

Parsed values

```
<mods:extension>  
  <etd:degree>  
    <etd:level/>  
    <etd:discipline/>  
  </etd:degree>  
</mods:name>
```



# Migrating ETDs from Dublin Core to MODS

## Core Description

Parsed values

```
<keywords>  
  <keyword/>  
</keywords>
```

## ***Full Description***

Text string

```
<mods:note  
  displayLabel="Keyword"/>
```



# Migrating ETDs from Dublin Core to MODS

## Core Description

Publication Date Only

`<publication-date/>`

## ***Full Description***

Publication Statement

```
<mods:originInfo>  
  <mods:place>  
    <mods:placeTerm/>  
  </mods:place>  
  <mods:dateIssued/>  
  <mods:publisher/>  
  <mods:issuance/>  
</mods:titleInfo>
```

Includes other publication info

# Migrating ETDs from Dublin Core to MODS

## Core Description

Parsed first & middle  
Uncontrolled form of name

```
<author>  
  <institution/>  
  <lname/>  
  <fname/>  
  <mname/>  
  <suffix/>  
</author>
```

## *Full Description*

First & middle = given  
Controlled form of name

```
<mods:name type="personal">  
  <mods:namePart type="date"/>  
  <mods:namePart type="family"/>  
  <mods:namePart type="termsOfAddress"/>  
  <mods:namePart  
    type="given"/>  
  <mods:role>  
    <mods:roleTerm>author  
    </mods:roleTerm>  
  </mods:role>  
</mods:name>
```



# Migrating ETDs from Dublin Core to MODS

## Core Description

Text string

```
<fields>  
  <field name=  
    "department">  
    <value/>  
  </field>  
</fields>
```

## ***Full Description***

Controlled form of name

```
<mods:name  
  type="corporate">  
  <mods:namePart/>  
  <mods:role>  
    <mods:roleTerm>  
      degree-granting department  
    </mods:roleTerm>  
  </mods:role>  
</mods:name>
```

# Migrating ETDs from Dublin Core to MODS

## Core Description

HTML

styling & references

`<abstract/>`

---

```
<fields>
  <field
    name="advisor1">
    <value/>
  </field>
</fields>
```

## *Full Description*

UTF-8

character encoding

`<abstract/>`

---

```
<mods:note
  displayLabel=
  "Advisory Committee"/>
```



## Keyword Search

Search retrieves record for display.

Name authority control is not necessary.

## ***Post-Search Faceting***

Search retrieves record for display *and* "Narrow Results By" faceting.

Name authority control *is* necessary.

## Goals

- ◆ RDA and ETD-MS standards
- ◆ Accurate title transcription
- ◆ Authority control of author names
- ◆ Library of Congress Subject Headings
- ◆ Notes for:
  - ▶ Submitted Statement
  - ▶ Degree Awarded
  - ▶ Advisory Committee



## Survey

Which of these XML technologies are you familiar with? (Multiple answers allowed)

- ◆ XSLT
- ◆ XPath
- ◆ XProc
- ◆ XQuery
- ◆ None of the above

## eXtensible Stylesheet Language Transformation

### XSLT 1.0

- ◆ “Push” to navigate the XML node tree using `<apply-template>`
- ◆ Designed for schema-to-schema crosswalks

### XSLT 2.0

- ◆ “Pull” to navigate the XML node tree using XPath functions
- ◆ Enhanced capability for:
  - ▶ Regular expressions
  - ▶ Conditional statements
  - ▶ Iterative expressions



PDF Data



## Generating PDF data

### 1. Process PDFs

- 1.1. Strip Bepress title page
- 1.2. Extract pages 1-2
- 1.3. Save PDF as XML (pdf2xml)

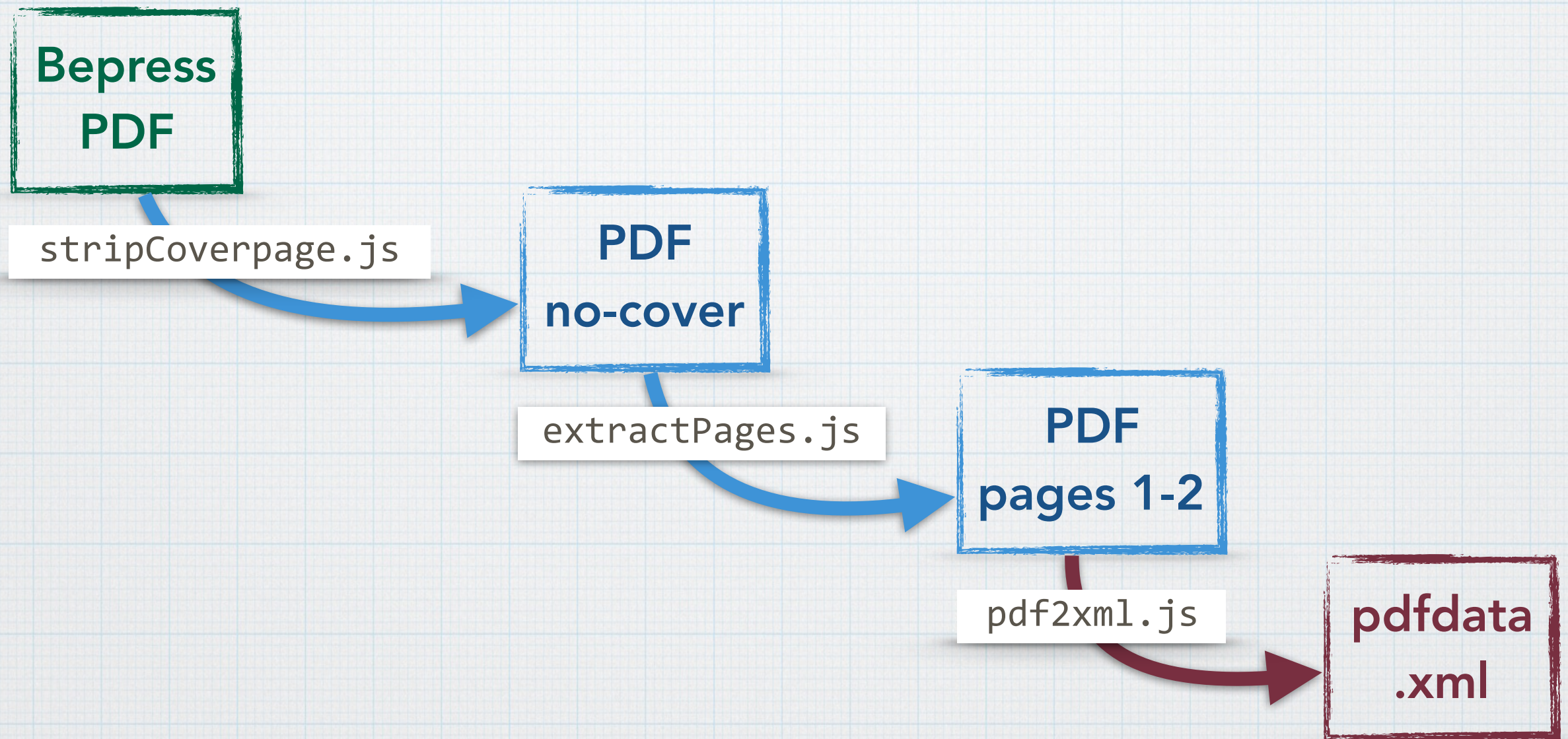
### 2. Create pdfdata.xml

- 2.1. Run XProc
- 2.2. Correct pdf2xml as needed
- 2.3. Review and correct final result



# Migrating ETDs from Dublin Core to MODS

## JavaScript Adobe Acrobat Wizard





# Migrating ETDs from Dublin Core to MODS

## JavaScript **Adobe Acrobat Wizard**

### removePage.js

```
for (var n = 0; n < this.getPagNumWords(0); n++) {  
    if (this.getPageNthWord(0, n) == "DigiNole") {  
        this.deletePages(0);  
        break;  
    }  
}
```



# Migrating ETDs from Dublin Core to MODS

## JavaScript **Adobe Acrobat Wizard**

### **extractPages.js**

```
var filename =  
  this.documentFileName.split(".").shift()+"_pages1-2.pdf";  
this.extractPages({  
  nStart: 0, nEnd: 1, cPath : filename  
});
```

# Migrating ETDs from Dublin Core to MODS

## JavaScript **Adobe Acrobat Wizard**

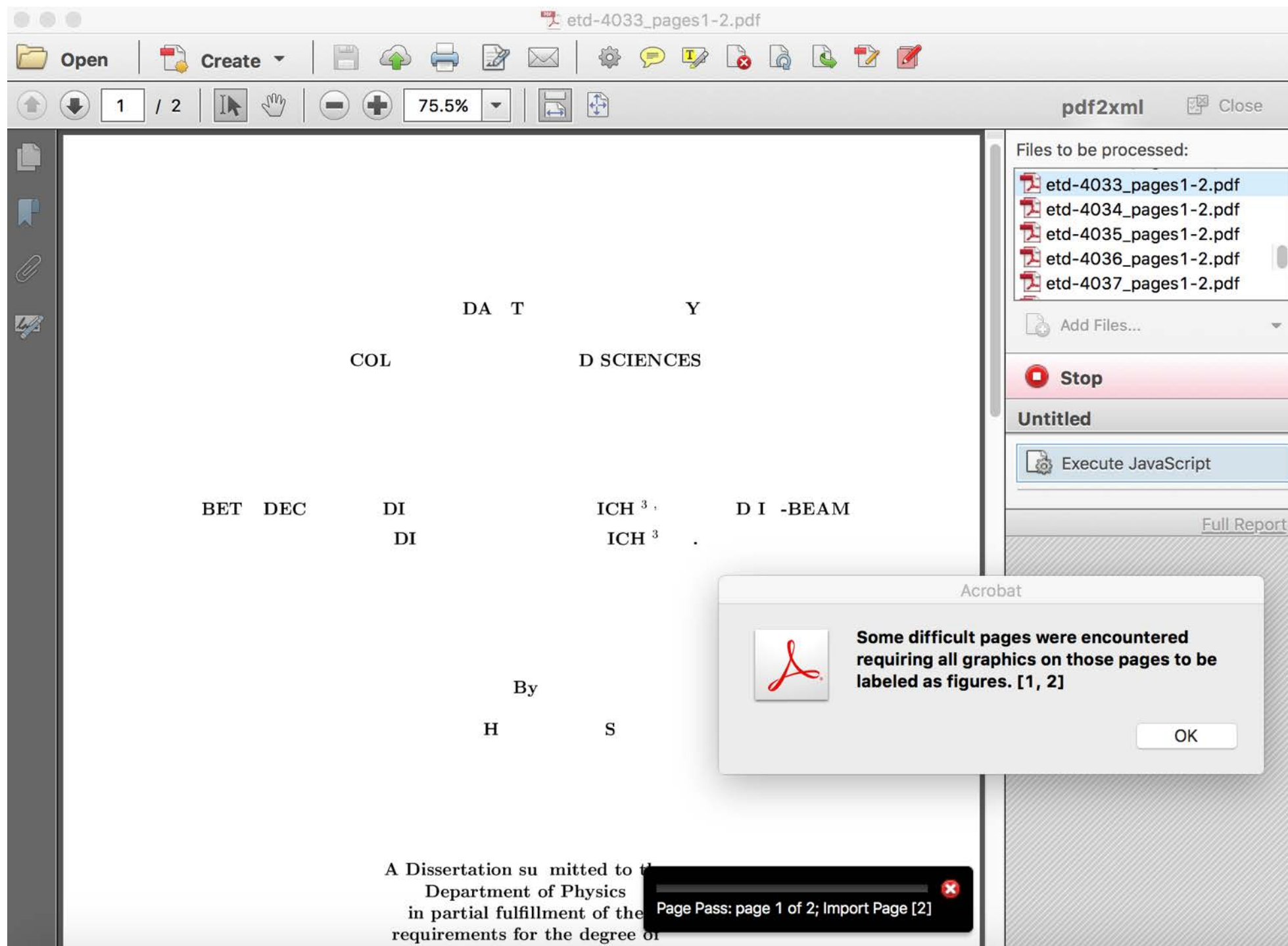
### pdf2xml.js

```
var filename =  
  this.documentFileName.split("_pages1-2.pdf").shift()  
  + "_pdfdata.xml";  
this.saveAs(filename,  
  "com.adobe.acrobat.xml-1-00" nStart: 0,nEnd: 1, cPath  
  : filename);
```



# Migrating ETDs from Dublin Core to MODS

## Conversion Issues



# Migrating ETDs from Dublin Core to MODS

## pdf2xml

<TaggedPDF-doc>

[snip]

<Figure>

<ImageData src="images/etd-4033\_pdfdata\_img\_1.jpg"/>

i M R M R z l M a o d e d i n M M m d R l R R d d M M e a a n a d M d i M  
R o o M a d d R d e d e a R o i d r R e e n M a o e n M r M e n e M e d  
e i a b i n e P \_ R y t M \_ \_ e z d a a d r M o o d a r R  
a M b d R e R o o M a d d R d e \_ t d d d a M \_ e a d e e n e  
R E t n d o R e n M M m d R l R R d d M M i M R z l M a \_ t n o e e \_ m \_ M o o R e M m d  
R R R d d M M i M R y z M a i a \_ \_ m R \_ M M m d R R R d d M M i M R y z M  
a \_ e d a M R M \_ a M \_ m R b \_ M m d R R R d d M M i M R y z M a a a a d  
e M e n i a \_ \_ m R \_ M M m i d e R a M a d e a d R l M e t d m d e t i t o R b  
o \_ e d o M a i a m R o f M e e n M m d \_ \_ M M d e a d o r e e e \_ b  
R M e b M o t m m m m e  
m t m m r

</Figure>

</TaggedPDF-doc>

# Migrating ETDs from Dublin Core to MODS

## pdf2xml to pdfdata

<lines>

<number>etd-4033</number>

<line>i M R M R z l M a o d e d i n M M m d R l R R d d M M e a a n a d M d i M R  
o o M a d d R d e d e a R o i d r R e e n M a o e n M r M e n e M e d e i a b i n  
e P \_ R y t M \_ \_ e z d a a d r M o o d a r R a M b d R e R o o M a d d R d  
e \_ t d d d a M \_ e a d e e n e R E t n d o R e n M M m d R l R R d d M M i M R z l M  
a \_ t n o e e \_ m \_ M o o R e M m d R R R d d M M i M R y z M a i a \_ \_ m R \_ M M m d  
R R R d d M M i M R y z M a \_ e d a M R M \_ a M \_ m R b \_ M m d R R R d d M M i  
M R y z M a a a a d e M e n i a \_ \_ m R \_ M M m i d e R a M a d e a d R l M e t d m d e t  
i t o R b o \_ e d o M a i a m R o f M e e n M m d \_ \_ M M d e a d o r e e e \_ b R M  
e b M o t m m m e m t m m r</line>

</lines>



# Migrating ETDs from Dublin Core to MODS

## Solution - Enter Manually

<TaggedPDF-doc>

<P>header</P>

<P>FLORIDA STATE UNIVERSITY</P>

<P>College of [COLLEGE]</P>

<P>[TITLE]</P>

<P>[AUTHOR]</P>

<P>A [DOC\_TYPE] submitted to the Department of [NAME] in partial fulfillment of the requirements for the degree of [DEGREE].</P>

<P>Degree Awarded: [SEASON] Semester, [YEAR]</P>

<P>The members of the Committee approve the [DOC\_TYPE] of [AUTHOR] defended on [DATE].</P>

<P>[NAME] Professor Directing [DOC\_TYPE]</P>

<P>[NAME] Outside Committee Member</P>

<P>[NAME] Committee Member</P>

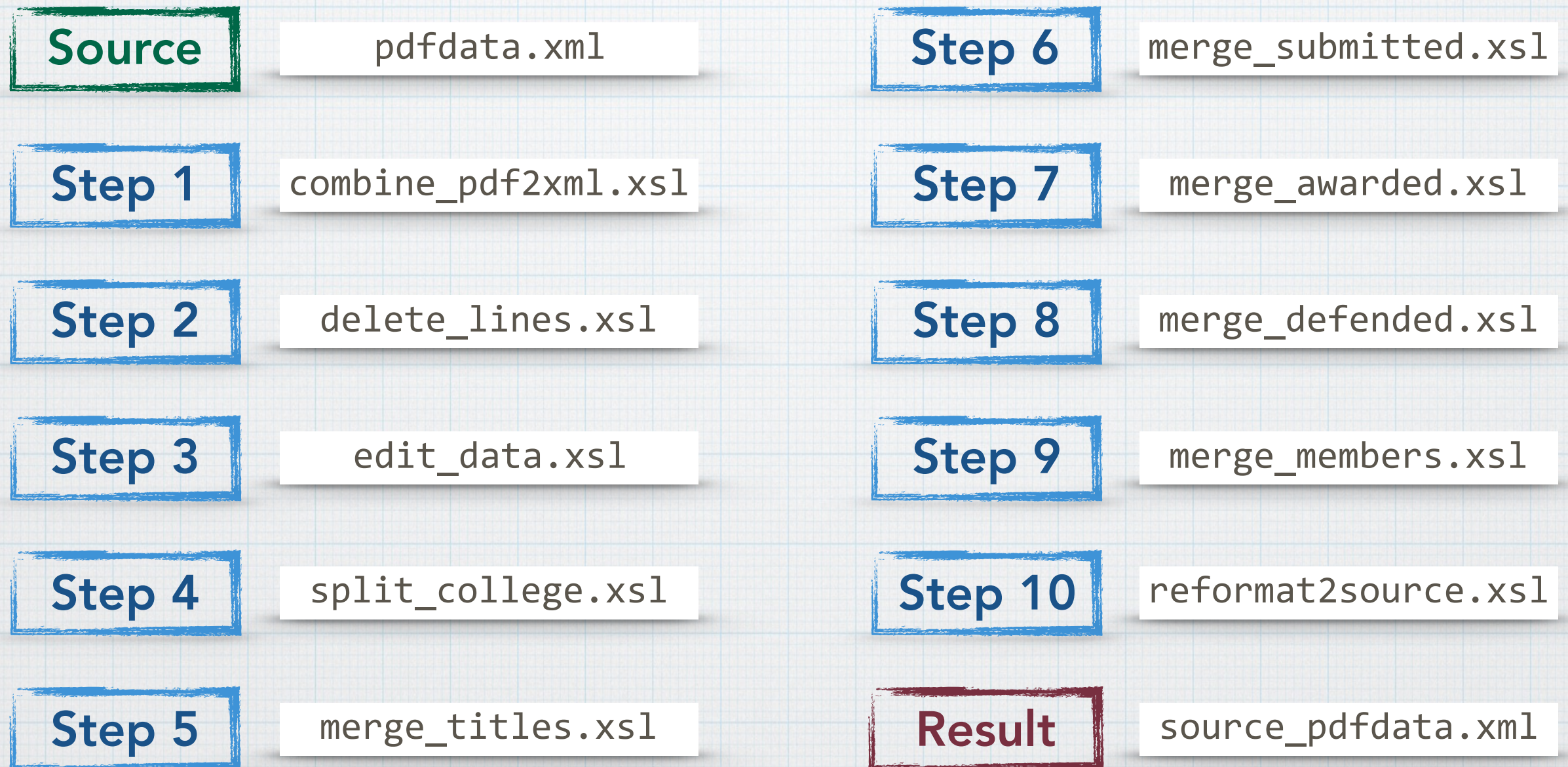
<P>[NAME] Committee Member</P>

</TaggedPDF-doc>



# Migrating ETDs from Dublin Core to MODS

## XProc





# Migrating ETDs from Dublin Core to MODS

## pdf2xml

<TaggedPDF-doc>

[snip]

<P>THE FLORIDA STATE UNIVERSITY </P>

<P>COLLEGE OF ARTS AND SCIENCES </P>

<P>“HOW WE GOT OVAH”: AFROCENTRIC SPIRITUALITY IN BLACK ARTS [...] </P>

<P>By </P>

<P>DARA TAFAKARI GREEN </P>

<P>A Thesis submitted to the Department of English [...] </P>

<P>Degree Awarded: Summer Semester, 2007 </P>

<P>The members of the Committee approve the Thesis of [...] </P>

<P>\_\_\_\_\_ </P>

<P>Jerrilyn McGregory </P>

<P>Professor Directing Thesis </P>

<P>Maxine Montgomery </P>

<P>Committee Member </P>

<P>Dennis Moore </P>

<P>Committee Member </P>

<P>Approved: </P>

<P>Nancy B. Warren Director of Graduate Studies </P>

<P>The Office of Graduate Studies has [...] committee members. </P>

<P>ii </P>

</TaggedPDF-doc>



# Migrating ETDs from Dublin Core to MODS

## Step 1 Result

```
<lines>
  <number>etd/4007</number>
  <line>THE FLORIDA STATE UNIVERSITY</line>
  <line>COLLEGE OF ARTS AND SCIENCES</line>
  <line>“HOW WE GOT OVAH”: AFROCENTRIC SPIRITUALITY IN [...]</line>
  <line>By</line>
  <line>DARA TAFAKARI GREEN</line>
  <line>A Thesis submitted to the Department of [...]</line>
  <line>Degree Awarded: Summer Semester, 2007</line>
  <line>The members of the Committee approve the [...]</line>
  <line>_____</line>
  <line>Jerrilyn McGregory</line>
  <line>Professor Directing Thesis</line>
  <line>Maxine Montgomery</line>
  <line>Committee Member</line>
  <line>Dennis Moore</line>
  <line>Committee Member</line>
  <line>Approved:</line>
  <line>Nancy B. Warren Director of Graduate Studies</line>
  <line>The Office of Graduate Studies [...] committee members.</line>
  <line>ii</line>
</lines>
```

# Migrating ETDs from Dublin Core to MODS

## Step 2 Process

```
<lines>
  <number>etd/4007</number>
  <line>THE FLORIDA STATE UNIVERSITY</line>
  <line>COLLEGE OF ARTS AND SCIENCES</line>
  <line>“HOW WE GOT OVAH”: AFROCENTRIC SPIRITUALITY IN [...]</line>
  <del><line>By</line>
  <line>DARA TAFAKARI GREEN</line>
  <line>A Thesis submitted to the Department of [...]</line>
  <line>Degree Awarded: Summer Semester, 2007</line>
  <line>The members of the Committee approve the [...]</line>
  <del><line>_____</line>
  <line>Jerrilyn McGregory</line>
  <line>Professor Directing Thesis</line>
  <line>Maxine Montgomery</line>
  <line>Committee Member</line>
  <line>Dennis Moore</line>
  <line>Committee Member</line>
  <del><line>Approved:</line>
  <line>Nancy B. Warren Director of Graduate Studies</line>
  <line>The Office of Graduate Studies [...] committee members.</line>
  <del><line>ii</line>
</lines>
```

# Migrating ETDs from Dublin Core to MODS

## Step 2 Result

```
<lines>
  <number>etd/4007</number>
  <line>THE FLORIDA STATE UNIVERSITY</line>
  <line>COLLEGE OF ARTS AND SCIENCES</line>
  <line>“HOW WE GOT OVAH”: AFROCENTRIC SPIRITUALITY IN [...]</line>
  <line>DARA TAFAKARI GREEN</line>
  <line>A Thesis submitted to the Department of [...]</line>
  <line>Degree Awarded: Summer Semester, 2007</line>
  <line>The members of the Committee approve the [...]</line>
  <line>Jerrilyn McGregory</line>
  <line>Professor Directing Thesis</line>
  <line>Maxine Montgomery</line>
  <line>Committee Member</line>
  <line>Dennis Moore</line>
  <line>Committee Member</line>
</lines>
```



# Migrating ETDs from Dublin Core to MODS

## Final Result

```
<record>
  <number>etd/4007</number>
  <college>College of Arts and Sciences</college>
  <title>“How We Got Ovah”: Afrocentric Spirituality in [...]</title>
  <author>Dara Tafakari Green</author>
  <submitted>A Thesis submitted to the Department of English [...]</submitted>
  <awarded>Degree Awarded: Summer Semester, 2007</awarded>
  <defense>June 25, 2007</defense>
  <committee>
    <member>
      <name>Jerrilyn McGregory</name>
      <role>Professor Directing Thesis</role>
    </member>
    <member>
      <name>Maxine Montgomery</name>
      <role>Committee Member</role>
    </member>
    <member>
      <name>Dennis Moore</name>
      <role>Committee Member</role>
    </member>
  </committee>
</record>
```

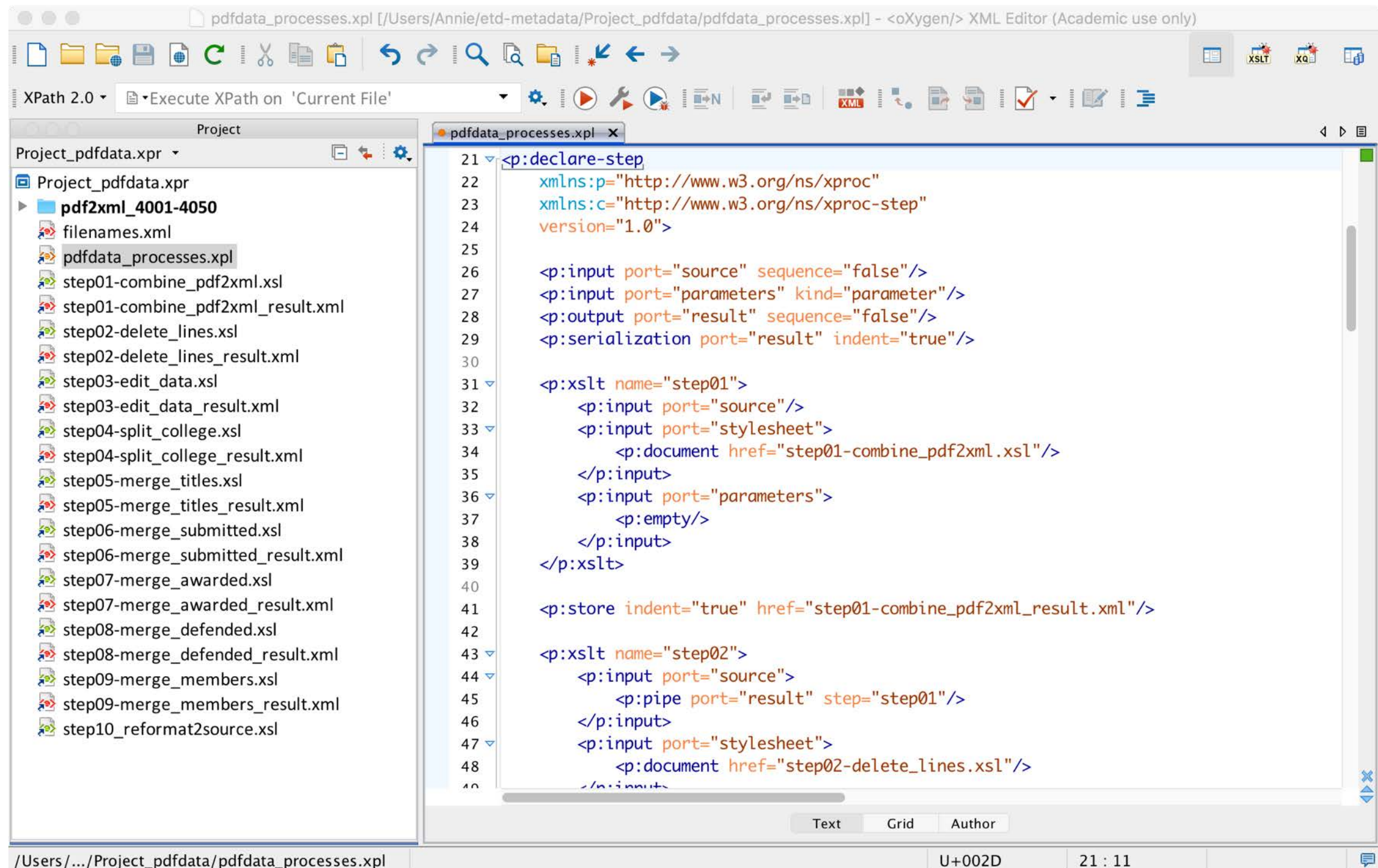


# Modular XSLT



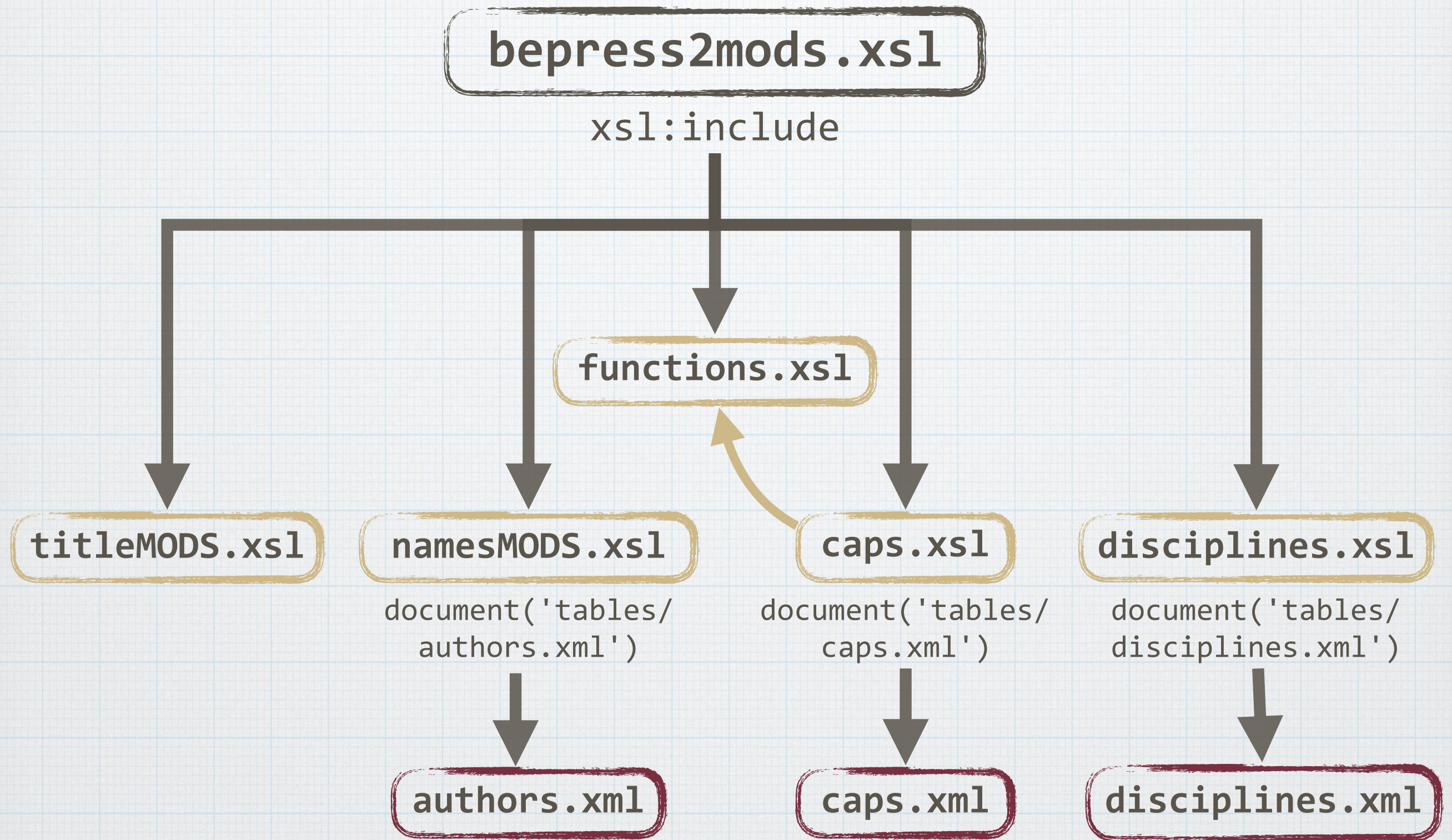
# Migrating ETDs from Dublin Core to MODS

## oXygen XML Editor





# Migrating ETDs from Dublin Core to MODS

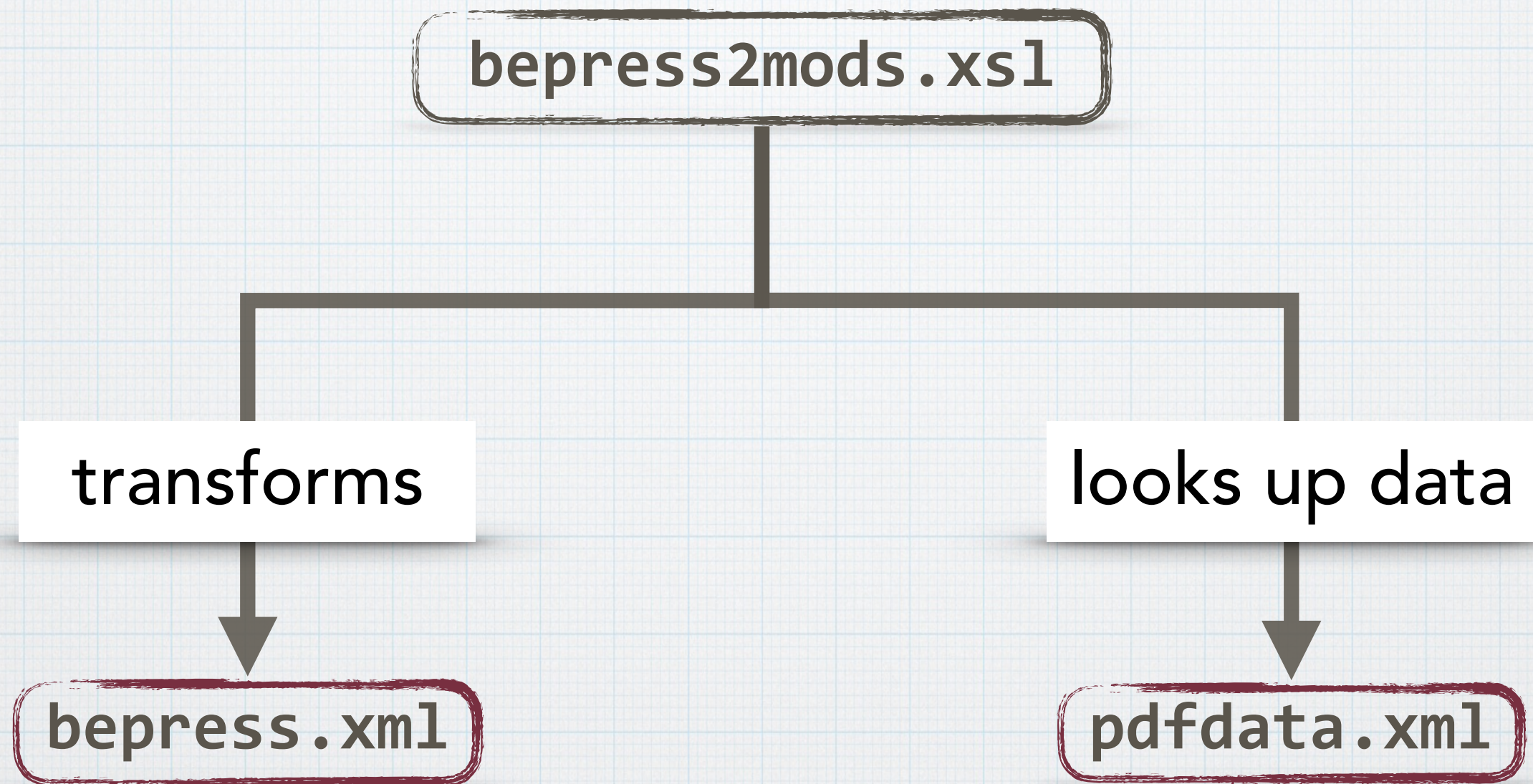




**Multiple XML**



# Migrating ETDs from Dublin Core to MODS





# Migrating ETDs from Dublin Core to MODS

## Match Point

### Bepress

```
<submission-path>8516  
</submission-path>
```

```
<xsl:variable name="bepress"  
select=  
"documents/document/label"/>
```

### pdfdata

```
<number>8516</number>
```

```
<xsl:variable name="pdfdata"  
select=  
"document('pdfdata.xml')  
/records/record"/>
```

```
select="$pdfdata/number"
```

## Match Point

**Bepress**

```
<submission-path>8516  
</submission-path>
```

**pdfdata**

```
<number>8516</number>
```

```
<xsl:variable name="title">  
  <xsl:for-each select="$bepress">  
    <xsl:value-of  
      select="$pdfdata[number=current()]/  
        *[(self::title)]"/>  
  </xsl:for-each>  
</xsl:variable>
```

**Is this process complex?**

**— A bit**

**Is it efficient?**

**— You bet!**



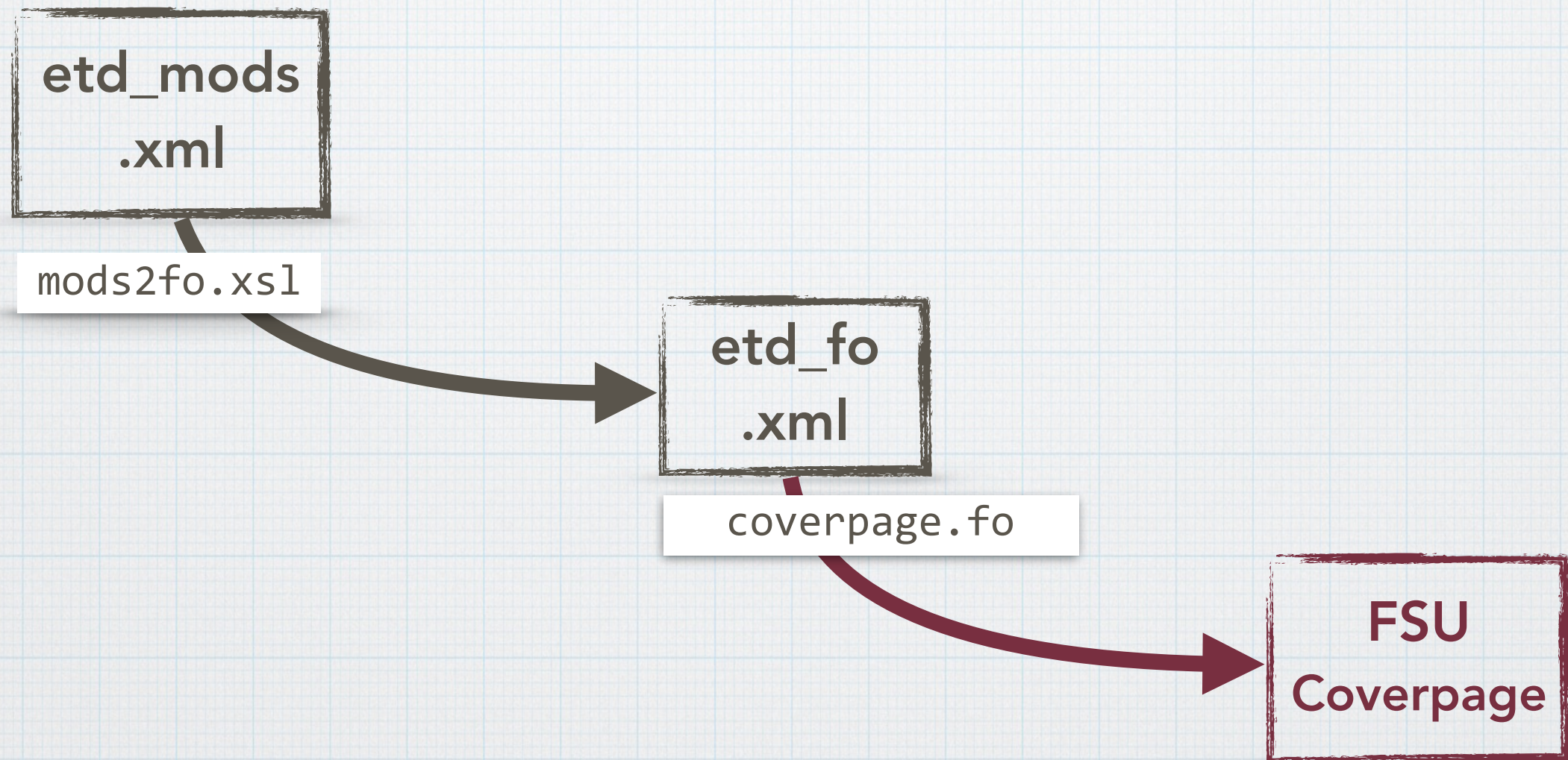
# Other Technologies



# Migrating ETDs from Dublin Core to MODS

## FO-XSL

for creating new cover page





# Migrating ETDs from Dublin Core to MODS

## Command Prompt for copying files

Start with

Source  
Folder

Create

Destination  
Folder

&

.TXT  
#PATH *not*  
to copy



## Command Prompt

for copying files

**Xcopy /I**

**SOURCE\_PATH**

**DESTINATION\_PATH**

**/Excludes:**

**TXT\_PATH**

## Command Prompt

for copying files

**Xcopy /I**

```
s:\bjbrown\diginole\source-data\etds\unsup-etd\fulltext  
\fulltext_1-1000\allpdfs0
```

```
s:\bjbrown\diginole\source-data\etds\unsup-etd\fulltext  
\fulltext_1-1000\badpdf0 /
```

**/Exclude:**

```
S:\bjbrown\diginole\source-data\etds\unsup-etd\fulltext  
\fulltext_1-1000\filelist0.txt
```



# Adaptations



**Bepress to MARC21**



## ETD Workflows at USF

**Pre-2015:** Student submits to ProQuest

- ◆ Batch processing through XSLT, shell scripting

**Spring 2015:** Student submits to ProQuest

- ◆ IR metadata was entered manually
- ◆ MARC records were entered manually

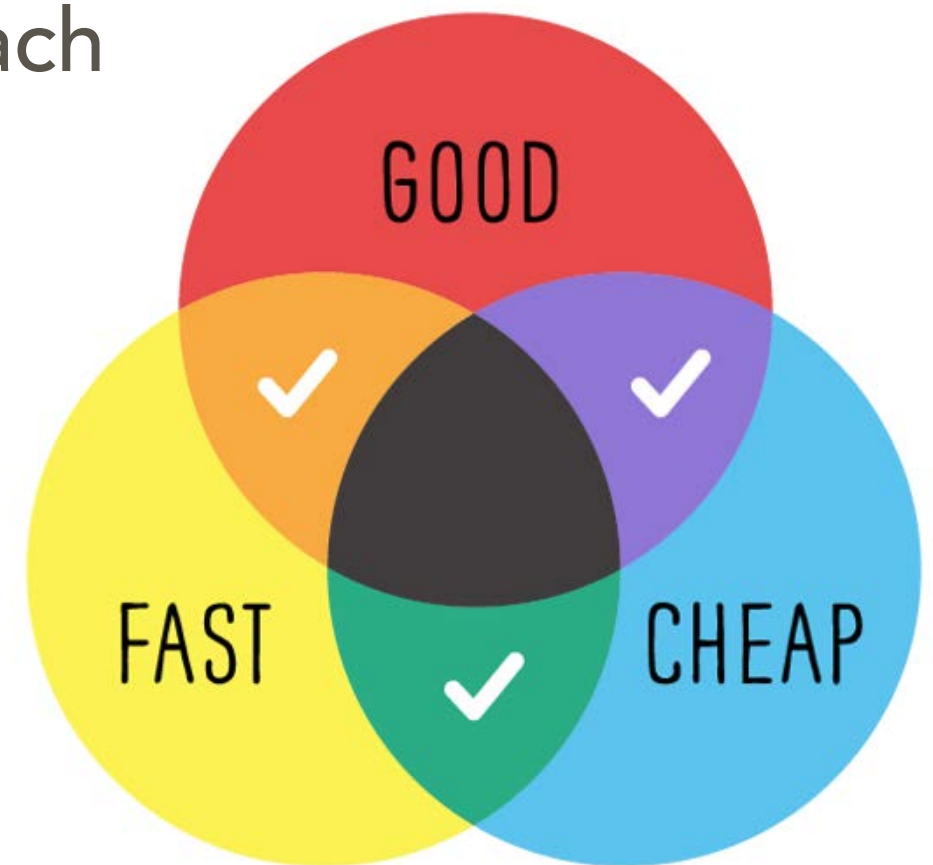
**Fall 2015-Present:** Student submits to ProQuest

- ◆ XSLT automation creates IR metadata
- ◆ XSLT automation creates MARC records



## Updating the Workflow

- ◆ “Fast, cheap, *and* good” approach
- ◆ Full level MARC21 records
- ◆ RDA and ETD-MS standards
- ◆ Authority control of authors
- ◆ Notes for:
  - ◆ Degree awarded
  - ◆ Committee members
- ◆ Library of Congress Subject Headings



## ETD Workflow Inspiration

### Florida State University

#### Annie Glerum

- ◆ Modular XML technologies for generating ETD metadata
- ◆ Efficiencies for Quality Control of Repurposed Metadata
- ◆ Adding XSLT to the Cataloger's Toolbox

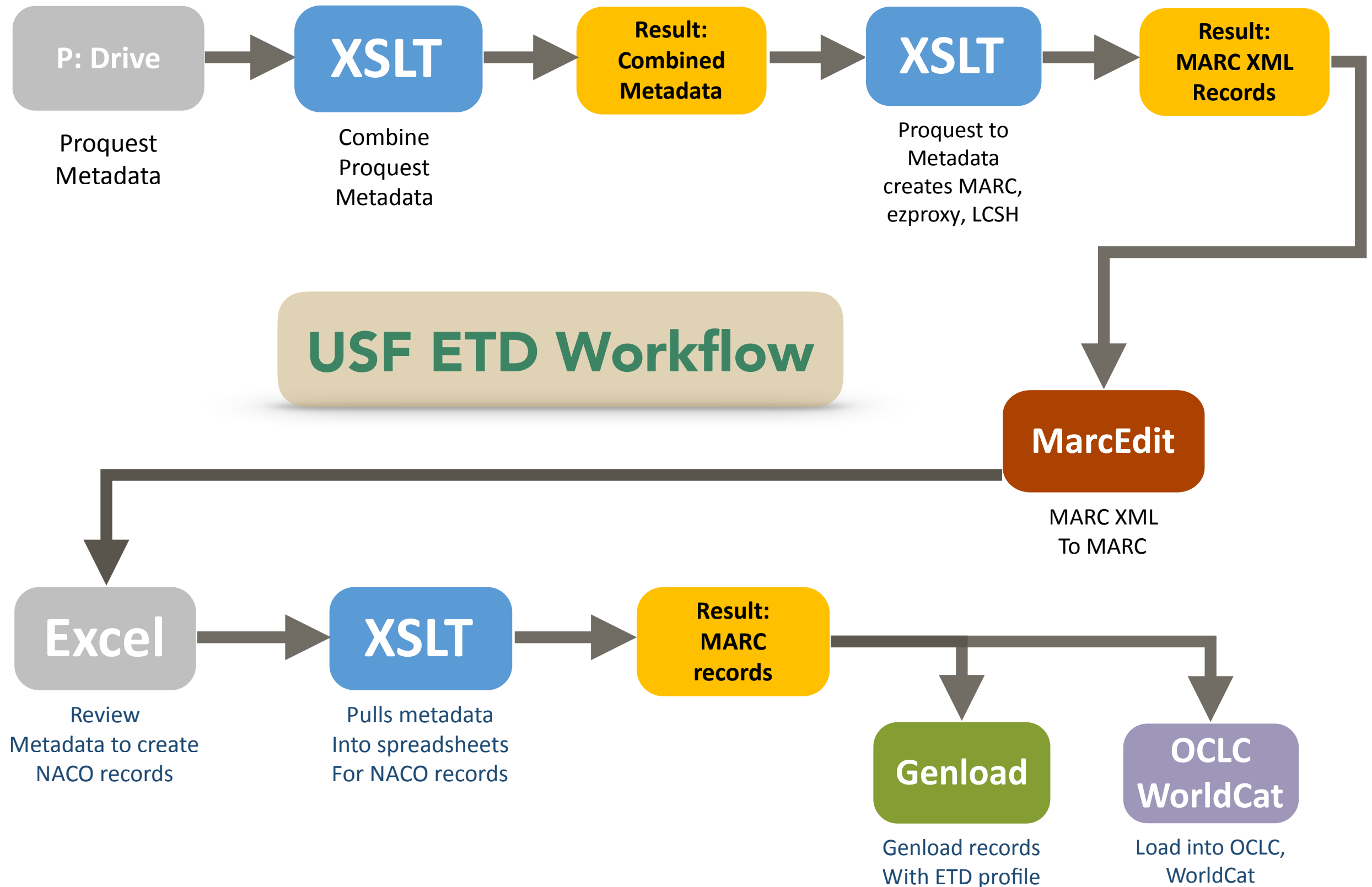
### University of Iowa

#### Shawn Averkamp & Joanna Lee

- ◆ Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository



# Migrating ETDs from Dublin Core to MODS



## Combine Metadata

Unzip ProQuest  
Metadata XML

Copy Filepaths

Combine ProQuest  
Metadata

Result:  
Master source file.xml

```
<xsl:template match="/">
```

```
  <xml>
```

```
    <xsl:copy-of select="document ('file:///u:/ETD_Project/  
2015-09-14/ original_xml/[filename].xml')"/>
```

```
  </xml>
```

```
  <xsl:result-document method="xml" href="combined-  
files.xml"/>
```

```
</xsl:template>
```



# Migrating ETDs from Dublin Core to MODS

## ProQuest to Bepress

ProQuest (source)	Bepress (output)
xml	documents
DISS_submission	document
DISS_description/DISS_title	title
DISS_authorship/DISS_author/DISS_name/DISS_surname	authors/author/lname
DISS_affiliation	authors/author/institution (University of South Florida)
DISS_content/DISS_abstract	abstract
DISS_description/DISS_dates/DISS_comp_date	publication-date
DISS_description/DISS_categorization/DISS_keyword	keywords/keyword
DISS_content/DISS_binary	fulltext-url
DISS_description/DISS_degree	fields/field/@name="degree_name"/value
DISS_description/DISS_institution/DISS_inst_contact	fields/field/@name="department"/value
DISS_description/DISS_advisor[1]/DISS_name/DISS_fname	fields/field/@name="advisor[1]"/value

## ProQuest to Bepress


```
<authors>
  <xsl:for-each select="DISS_authorship/DISS_author">
    <author xsi:type="individual">
      <institution>University of South Florida</institution>
      <lname>
        <xsl:variable name="lastname" select="DISS_name/DISS_surname"/>
        <xsl:variable name="hyphenlastname"
          select="replace($lastname, '-', '~')"/>
        <xsl:value-of
          select="replace(str:title-case(normalize-space($hyphenlastname)), '- ', '-')"/>
      />
    </lname>
  </for-each>
</authors>
```



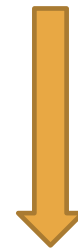
# Migrating ETDs from Dublin Core to MODS

## ProQuest to Bepress


```
<?xml version="1.0" encoding="ISO-8859-1"?>
<DISS_submission publishing_option="0" embargo_code="0" third_party_search="Y">
  <DISS_authorship>
    <DISS_author type="primary">
      <DISS_name>
        <DISS_surname>Smith</DISS_surname>
        <DISS_fname>Mary</DISS_fname>
        <DISS_middle>Ann</DISS_middle>
        <DISS_suffix/>
        <DISS_affiliation>University of South Florida</DISS_affiliation>
      </DISS_name>
    </DISS_author>
  </DISS_authorship>
</DISS_submission>
```



Proquest



```
<authors>
  <author xsi:type="individual">
    <institution>University of South Florida</institution>
    <lname>Smith</lname>
    <fname>Mary</fname>
    <mname>Ann</mname>
    <suffix/>
  </author>
</authors>
```



Bepress

# Migrating ETDs from Dublin Core to MODS

## Bepress to MARC

Bepress (source)	MARC (output)
documents	<marc:collection>
document	<marc:record>
title	<marc:datafield tag="245" ind1="1" ind2=" " >
authors/author/lname	<marc:datafield tag="100" ind1="1" ind2=" " >
abstract	<marc:datafield tag="520" ind1="3" ind2=" " >
publication-date	<marc:datafield tag="264" ind1=" " ind2="1" >
keywords/keyword	<marc:datafield tag="650" ind1=" " ind2="0" >
fulltext-url	<marc:datafield tag="856" ind1="4" ind2="0" >
fields/field/@name="degree_name"/value	<marc:datafield tag="502" ind1=" " ind2=" " >
fields/field/@name="advisor[1]"/value	<marc:datafield tag="700" ind1="1" ind2=" " >



# Migrating ETDs from Dublin Core to MODS

## Bepress to MARC

```
<marc:datafield tag="100" ind1="1" ind2=" " >
  <marc:subfield code="a">
    <xsl:variable name="fname" select="authors/author/fname"/>
    <xsl:variable name="mname" select="authors/author/mname"/>
    <xsl:variable name="lname" select="authors/author/lname"/>
    <xsl:choose>
      <xsl:when test="$mname!=''">
        <xsl:value-of select="concat($lname, ', ', $fname, ', ', $mname, ', ')" />
      </xsl:when>
      <xsl:otherwise>
        <xsl:value-of select="concat($lname, ', ', $fname, ', ')" />
      </xsl:otherwise>
    </xsl:choose>
  </marc:subfield>
  <marc:subfield code="e">author</marc:subfield>
</marc:datafield>
```

# Migrating ETDs from Dublin Core to MODS

## ProQuest to Bepress to MARC

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<DISS_submission publishing_option="0" embargo_code="0" third_party_search="Y">
  <DISS_authorship>
    <DISS_author type="primary">
      <DISS_name>
        <DISS_surname>Smith</DISS_surname>
        <DISS_fname>Mary</DISS_fname>
        <DISS_middle>Ann</DISS_middle>
        <DISS_suffix/>
        <DISS_affiliation>University of South Florida</DISS_affiliation>
      </DISS_name>
    </DISS_author>
  </DISS_authorship>
</DISS_submission>
```

Proquest



```
<authors>
  <author xsi:type="individual">
    <institution>University of South Florida</institution>
    <lname>Smith</lname>
    <fname>Mary</fname>
    <mname>Ann</mname>
    <suffix/>
  </author>
</authors>
```

Bepress

```
<marc:datafield tag="100" ind1="1" ind2=" " >
  <marc:subfield code="a">Smith, Mary Ann</marc:subfield>
  <marc:subfield code="e">author</marc:subfield>
</marc:datafield>
<marc:datafield tag="264" ind1=" " ind2="1">
  <marc:subfield code="a">Tampa, Florida :</marc:subfield>
  <marc:subfield code="b"> University of South Florida,</marc:subfield>
  <marc:subfield code="c">2015.</marc:subfield>
</marc:datafield>
```

MARC





## Bepress to MARC Result

One master XML file for all ETD MARC records

Convert XML to MARC in MarcEdit

Load into catalog through Genload

```
=LDR 04947ntm a2200301ii 4500
=006 m\ \ \ \ \ \ d \ \ \ \ \ \
=007 cr \ n \ \ \ \ \ \ \ \ \ \
=008 \ \ \ \ \ s 2015 \ \ \ \ \ xx \ \ \ \ \ sbm \ \ \ 000 \ 0 \ leng \ d
=040 \ \ $ a FHM $ beng $ erda $ c FHM
=100 1 \ $ a Bleck, Jennifer Rebecca $ e author
=245 10 $ a Exploring The Underlying Mechanisms Of Comorbid Adhd And Eating Disorders/ $ c by Jennifer Rebecca Bleck.
=264 1 \ $ a Tampa, Florida : $ b University of South Florida, $ c 2015.
=336 \ \ $ a text $ b txt $ 2 rda content
=337 \ \ $ a computer $ bc $ 2 rda media
=338 \ \ $ a online resource $ b cr $ 2 rda carrier
=500 \ \ $ a Thesis supervisor: DeBate, Rita R.
=500 \ \ $ a Thesis supervisor: Levin, Bruce L.
=502 \ \ $ b Ph. D. $ c University of South Florida, Tampa $ d 2015
```



# NACO List



## Automating NACO Authority Records

- ◆ “Fast, cheap, *and* good” approach
- ◆ Identify metadata for NACO authority records
- ◆ XSLT copies metadata into Excel
- ◆ Metadata Librarian creates NACO authorities



# Automating NACO Authority Records

```
<xsl:template name="hol" match="/">
  <root>
    <xsl:for-each select="*/marc:record">
      <xsl:variable name="author">
        <xsl:value-of select="marc:datafield[@tag='100']"/>
      </xsl:variable>
      <xsl:variable name="title">
        <xsl:value-of select="marc:datafield[@tag='245']"/>
      </xsl:variable>
      <xsl:variable name="pub">
        <xsl:value-of select="marc:datafield[@tag='264']"/>
      </xsl:variable>
      <xsl:variable name="degree">
        <xsl:value-of select="marc:datafield[@tag='502']"/>
      </xsl:variable>
      <xsl:variable name="url">
        <xsl:value-of select="marc:datafield[@tag='856']"/>
      </xsl:variable>
    </xsl:for-each>
  </root>
</xsl:template>
```



# Migrating ETDs from Dublin Core to MODS

## Result

### Metadata for NACO in XML

```
<rows>
  <column>Juhlin, Cory Michael,</column>
  <column>Developing A Compiler For A Regular Expression Based Policy Specification
    Language / by Cory Michael Juhlin.</column>
  <column>Tampa, Florida : University of South Florida, 2015.</column>
  <column>MSCS</column>
  <column>2015</column>
  <column>http://scholarcommons.usf.edu/etd/5885</column>
</rows>
<rows>
  <column>Klein, Alissa,</column>
  <column>Getting Ahead : Socio-Economic Mobility, Perceptions of Opportunity For Socio-Economic
    Mobility, And Attitudes Towards Public Assistance In The United States / by Alissa Klein.</column>
  <column>Tampa, Florida : University of South Florida, 2015.</column>
  <column>MA</column>
  <column>2015</column>
  <column>http://scholarcommons.usf.edu/etd/5886</column>
</rows>
<rows>
  <column>Lee, Tiffany T.,</column>
  <column>The Influence of Negative Information On Trust In Virtual Teams / by Tiffany T. Lee.</column>
  <column>Tampa, Florida : University of South Florida, 2015.</column>
  <column>MA</column>
  <column>2015</column>
  <column>http://scholarcommons.usf.edu/etd/5887</column>
</rows>
```

# Migrating ETDs from Dublin Core to MODS

## Result

Metadata for NACO in Excel

Metadata librarian creates author record

## Future Enhancements

Transform NACO ETD metadata result  
into NACO authority record

100 Author	245 Title	264 Publication	Degree	Year	URL
Juhlin, Cory Michael,	Developing A Compiler For A Regul	Tampa, Florida : University of South Florida, 2015.	MSCS	2015	<a href="http://scholarcommons.usf.edu/etd/5885">http://scholarcommons.usf.edu/etd/5885</a>
Klein, Alissa,	Getting Ahead : Socio-Economic M	Tampa, Florida : University of South Florida, 2015.	MA	2015	<a href="http://scholarcommons.usf.edu/etd/5886">http://scholarcommons.usf.edu/etd/5886</a>
Lee, Tiffany T.,	The Influence of Negative Informatic	Tampa, Florida : University of South Florida, 2015.	MA	2015	<a href="http://scholarcommons.usf.edu/etd/5887">http://scholarcommons.usf.edu/etd/5887</a>
Mejias De Pernia, Yolibeth,	Prediction of The Optimum Binder (	Tampa, Florida : University of South Florida, 2015.	PhD	2015	<a href="http://scholarcommons.usf.edu/etd/5888">http://scholarcommons.usf.edu/etd/5888</a>
Nugent, Patrick Stephen,	Tailoring The Pore Environment of M	Tampa, Florida : University of South Florida, 2015.	MS	2015	<a href="http://scholarcommons.usf.edu/etd/5889">http://scholarcommons.usf.edu/etd/5889</a>
Thompson, Erika L.,	The Influence of Relationship Statu	Tampa, Florida : University of South Florida, 2015.	PhD	2015	<a href="http://scholarcommons.usf.edu/etd/5890">http://scholarcommons.usf.edu/etd/5890</a>
De Paoli, Federico,	Measuring Polydimethylsiloxane (p	Tampa, Florida : University of South Florida, 2015.	MSME	2015	<a href="http://scholarcommons.usf.edu/etd/5881">http://scholarcommons.usf.edu/etd/5881</a>
Hergert, Randall J.,	Saharan Air Layer Dust Loading : E	Tampa, Florida : University of South Florida, 2015.	MS	2015	<a href="http://scholarcommons.usf.edu/etd/5882">http://scholarcommons.usf.edu/etd/5882</a>
Hinds, Gregory Richard,	High-Solids Anaerobic Digestion of	Tampa, Florida : University of South Florida, 2015.	MSEV	2015	<a href="http://scholarcommons.usf.edu/etd/5883">http://scholarcommons.usf.edu/etd/5883</a>
Hutchinson, Derek Michael,	Employee Retaliation Against Abu	Tampa, Florida : University of South Florida, 2015.	MA	2015	<a href="http://scholarcommons.usf.edu/etd/5884">http://scholarcommons.usf.edu/etd/5884</a>



# Migrating ETDs from Dublin Core to MODS

## Acronyms

- ▶ ETD-MS = Electronic Theses and Dissertations - Metadata Standard
- ▶ ETD = Electronic Theses and Dissertations
- ▶ FSU = Florida State University
- ▶ HTML = HyperText Markup Language
- ▶ MODS = Metadata Object Description Schema
- ▶ NACO = Name Authority Cooperative
- ▶ PDF = Portable Document Format
- ▶ RDA = Resource Description and Access
- ▶ USF = University of South Florida
- ▶ UTF-8 = Unicode Transformation Format - 8-bit
- ▶ VIAF = Virtual International Authority File
- ▶ XML = eXtensible Markup Language
- ▶ XProc = XML Pipeline
- ▶ XSLT = eXtensible Stylesheet Language Transformation



# Migrating ETDs from Dublin Core to MODS

## Resources

- ▶ Averkamp, Shawn and Joanna Lee. Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository. *Code4Lib Journal*. Issue 7, 2009-06-26. <http://journal.code4lib.org/articles/1647>
- ▶ DigiNole: FSU's Digital Repository. <http://diginole.lib.fsu.edu>
- ▶ Dublin Core® Metadata Initiative. DCMI Home. <http://www.w3.org/TR/2010/REC-xproc-20100511/>
- ▶ Glerum, Annie. Adding XSLT to the Cataloger's Toolbox: Efficiencies for Transforming and Analyzing Bibliographic Data. <http://myweb.fsu.edu/aglerum/assets/Adding-XSLT-to-the-catalogers-toolbox.pdf>
- ▶ Glerum, Annie. Efficiencies for Quality Control of Repurposed Metadata. Poster for USetdA 2014. [http://purl.flvc.org/fsu/fd/FSU\\_migr\\_library\\_faculty\\_publications-0014](http://purl.flvc.org/fsu/fd/FSU_migr_library_faculty_publications-0014)
- ▶ Islandora Website. <http://islandora.ca/>



# Migrating ETDs from Dublin Core to MODS

## Resources (cont.)

- ▶ Library of Congress. Metadata Object Description Schema: MODS.  
<http://www.loc.gov/standards/mods/>
- ▶ Networked Digital Library of Theses and Dissertations. ETD-MS v1.1: an Interoperability Metadata Standard for Electronic Theses and Dissertations. <https://sites.google.com/a/ndltd.org/ndltd/standards/metadata#>
- ▶ oXygen XML Editor. <https://www.oxygenxml.com>
- ▶ RDA Toolkit. <http://www.rdatoolkit.org>
- ▶ RDA Registry. <http://www.rdaregistry.info>
- ▶ World Wide Web Consortium. XProc: An XML Pipeline Language: W3C Recommendation 11 May 2010. <http://www.w3.org/TR/xproc/>
- ▶ World Wide Web Consortium. XSL Transformations (XSLT) Version 2.0: W3C Recommendation 23 January 2007. <http://www.w3.org/TR/xslt20/>



# Questions?

---

*Contact Us!*

**Annie Glerum**

Head of Complex Cataloging  
Florida State University

[aglerum@fsu.edu](mailto:aglerum@fsu.edu)

<https://github.com/aglerum>

**Dominique Bortmas**

Metadata Librarian  
University of South Florida

[dbortmas@usf.edu](mailto:dbortmas@usf.edu)

<https://github.com/dbortmas12>