

6-29-2014

Metadata migration to Islandora: Is there an easy way?

Sai Deng

University of Central Florida, sai.deng@ucf.edu

Lee Dotson

University of Central Florida, lee.dotson@ucf.edu



Part of the [Cataloging and Metadata Commons](#)

Find similar works at: <https://stars.library.ucf.edu/ucfscholar>

University of Central Florida Libraries <http://library.ucf.edu>

This Conference Presentation is brought to you for free and open access by STARS. It has been accepted for inclusion in Faculty Scholarship and Creative Works by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

Original Citation

Deng, S. (2014). Metadata migration to Islandora: Is there an easy way? Metadata Interest Group Meeting, ALA Annual Conference 2014. Las Vegas, NV. June 29, 2014.

Metadata Interest Group Meeting, ALA Annual 2014

Metadata Migration to Islandora: Is There An Easy Way?



Sai Deng, Lee Dotson

University of Central Florida Libraries

What will be covered?

- A Fast Intro to MODS
- How does Islandora ingest data
- UCF Collections and Metadata Migration to Islandora
 - DC-MODS-MARC Transformation Challenges
 - Options to Improve Metadata: XSLT Stylesheet, MODS Editing
 - The Tool: Notepad++
 - Pattern Identification and MODS Editing
 - Issues to Think About
- Final thoughts

A Fast Intro to MODS

- **MODS: Metadata Object Description Schema** (<http://www.loc.gov/standards/mods/>)
- **MODS Elements and Attributes (Version 3.4)**
(<http://www.loc.gov/standards/mods/mods-outline-3-4.html>)
 - **Root Elements:** mods, modsCollection
 - **Top level elements:** titleInfo, name, typeOfResource, genre, originInfo, language, physicalDescription, abstract, tableOfContents, targetAudience, note, subject, classification, relatedItem, identifier, location, accessCondition, part, extension, recordInfo
 - **Attributes and subelements:** for example, name has attributes: ID, type, authority etc. and subelements: namePart, affiliation, role etc.
- **Conversion**
 - **LC MODS mappings**
 - DC to MODS 3.2/3.3/3.4 (<http://www.loc.gov/standards/mods/dcsimple-mods.html>)
 - MARC to MODS 3.4 (<http://www.loc.gov/standards/mods/mods-mapping-3-4.html>) (3.5 available)
 - MODS 3.4 to MARC (<http://www.loc.gov/standards/mods/v3/mods2marc-mapping.html>)
 - **LC MODS XSLT 1.0 Stylesheets (2.0 available)**
 - DC to MODS 3.4 (http://www.loc.gov/standards/mods/v3/DC_MODS3-4_XSLT1-0.xsl)
 - MARCXML to MODS 3.4 (<http://www.loc.gov/standards/mods/v3/MARC21slim2MODS3-4.xsl>)
 - MODS 3.4 to MARCXML (http://www.loc.gov/standards/mods/v3/MODS3-4_MARC21slim_XSLT1-0.xsl)

Islandora Migration @ UCF Libraries

- **Islandora** is an open source Digital Asset Management System (DAMS) based on FedoraCommons (as base repository), Drupal (as front-end CMS) and solr (as discovery application).
- **Islandora FLVC (Florida Virtual Campus)**
(https://islandora.pubwiki.fcla.edu/wiki/index.php/Main_Page)
- FL-Islandora is the FLVC instance hosted for the state universities and state colleges in a collaborative environment. All decisions are made collaboratively.
- **Islandora Collection and Metadata Migration @ UCF**
 - **DigiTool collections:** Political & Rights Issues & Social Movements (PRISM, available in DC format), Florida Heritage (MARCXML available)
 - **CONTENTdm collections:** University Archives, Civil War, Harrison Buzz Price, Theses and Dissertations, UCF Community Veterans History... (in CONTENTdm DC format)
 - How to prepare and work with metadata in this process?

FL-Islandora Ingest

- **Online ingest: choose content model and metadata form in Islandora GUI.**
 - **Content models:**
 - Unitary (one primary file, e.g. PDF files, single images);
 - Compound (child objects and parent objects, for any type of related objects to display together; parent objects containing metadata only);
 - Paged (hierarchical, for books, newspapers, can contain, e.g. newspaper, issue, page)
 - The metadata form is created by FLVC;
 - Librarians can manually input in the form; the form can pre-populate data through uploading a MARCXML file or from a template.
- **Batch import (Zip file importer): ingest a batch of objects through the online interface.**
 - Need to prepare zip file (zip the MODS metadata records and content files together, has strict rules)

FL-Islandora Ingest (Cont.)

- **Offline ingest: FTP content to server and program handles ingest.**
 - MODS record, content files and instructions called "manifest" packaged together; program handles the FTPed content;
 - Content can originally be from other systems, e.g. DigiTool, CONTENTdm. The ingested MODS files can be transformed from DC or MARCXML metadata;
 - DC to MODS, MARC to MODS XSLT stylesheets are involved in the conversion;
 - Offline ingest is FLVC programming, and it is not completed yet.
- **FL-Islandora has been customized by FLVC and it differs from a standard Islandora installation.**
- **Read more at:** https://islandora.pubwiki.fcla.edu/wiki/index.php/Main_Page
(Find and click "[FL-Islandora Documentation](#)" Google doc)

FL-Islandora Ingest (Cont.)

○ Final display and results:

- Brief and full record display (with file/page/image);
- MARC display (only to logged in users);
- Provide downloadable MARCXML.

○ FLVC Stylesheets:

- FLVC DC to MODS
(https://github.com/FLVC/FLVC_Islandora_XSLTs/blob/master/DC2MODS_DigiTool.xsl)
- FLVC MARC to MODS
(https://github.com/FLVC/FLVC_Islandora_XSLTs/blob/master/MARC21slim2MODS3-4_FLVC.xsl)
- FLVC MODS to MARC
(https://github.com/FLVC/FLVC_Islandora_XSLTs/blob/master/MODS3-4_MARC21slim_XSLT1-0_FLVC.xsl)
- MARC to HTML XSLT (for web display)

DC-MODS-MARC Transformation Challenges

- **DC-MODS-MARC transformation presents great challenges:**
 - Mapping is never perfect, or adequate (not in both ways).
 - The source schema (DC) has a very generic data representation;
 - DC: Only 15 elements. Has Qualifiers (but no "official" QDC-MODS stylesheet? Difficult to have one too)
 - The intermediate schema (MODS) cannot get enough values from DC for elements, sub-elements and attributes;
 - MODS: 20 top level elements, many sub-elements and attributes, more granular than DC
 - The target schema (MARC) does not really fit into the non-ILS type digital library environment? To keep it to enable better data sharing with an ILS?!
 - MARC 21: More granular than MODS, can have multiple MARC elements for a single MODS element

DC-MODS-MARC Transformation Challenges (Cont.)

○ Data ambiguity in DC-MODS mapping, for example,

DC	MODS	Notes
Title	<u><titleInfo><title></u>	Subtitle undistinguished. Non-filling characters, part name and number unmarked. Title type unclear: alternative, uniform...
Creator	<u><name><namePart><role><roleTerm type=text">creator...</u>	Creator type unclear: personal, corporate, conference. Sub-elements value unclear: <affiliation>, <role>, <description>...
Contributor	<u><name><namePart></u>	Name type unclear. Sub-elements value unclear.

○ Data ambiguity in MODS-MARC mapping, for example,

MODS	MARC	Notes
<u><title> with no <titleInfo> type attribute</u> <u><subTitle>...</u>	245 \$a with ind1=1 245 \$b	Title types (alternative, translated, uniform...) undistinguished; Unable to distinguish subtitle (if source data is from DC).
<u><name> with no type attribute</u>	720 ind1=blank ind2=blank (uncontrolled)	Authority control is not specified (when names are controlled). Cannot distinguish 1XX/7XX.
<u><name> with type="personal" roleterm "creator"</u>	100 ind1=1 ind2=blank	Not able to get name types from DC. Type value can be added. However, DC can have 1 creator, 2 or more creators in one record; when mapped to MARC, 100/700 can be problematic (unless we only take the first creator as the main author and define it in the stylesheet).
<u><name> with type="personal," no roleterm "creator"</u>	700 ind1=1 ind2=blank	The main author or all author names might not have the roleterm "creator." When mapping to MARC, the 100/700 relationship is not clear.

DC-MODS-MARC Transformation Challenges (Cont.)

○ Element relationship (DC-MODS-MARC conversion)

- Relationship between different authors: can not distinguish between 1XX(main entry, such as 100 Personal Name, 110 Corporate Name, 111 Meeting Name), and 7XX(added entry, such as 700, 710, 711);
- Author and title relationship: cannot generate correct indicators for 1XX and 245 (title) based on whether the author is the main entry;
- Subject relationships: unable to divide geographic, temporal, topical and genre subdivisions.

○ Local elements



- Many local elements for the collections (to be discussed. e.g. for University Archives, Civil War, Harrison Buzz Price, UCF Community Veterans History...)
- Can add local elements extensions, but it can cause difficulty in data sharing across collections and systems.

DC-MODS-MARC Transformation Challenges (Cont.)

- The LC DC-MODS and MODS-MARC stylesheets are impossible to produce adequate or desirable output (in nature), but what can be done to improve the results?
 - On the consortia level
 - DC-MODS
 - PURL (check for PURL, add one if none exist) (FLVC implemented)
 - Local elements: DigiTool pid (FLVC implemented)
 - ETD elements (to be added as a MODS extension)
 - MODS-MARC: Many changes have already been made to the LC stylesheet by FLVC before the UCF library started its migration, e.g.
 - Map MODS issuance to MARC 250, map typeOfResource to 998 77 \$b (so will be Mango format facet), map sublocation to 852 \$b, punctuate 260 and all main and added entry fields.
 - See: https://github.com/FLVC/FLVC_Islandora_XSLTs/blob/master/MODS3-4_MARC21slim_XSLT1-0_FLVC.xsl
 - Read more and check updates at: https://islandora.pubwiki.fcla.edu/wiki/index.php/Main_Page
 - Individual libraries can have their local stylesheets.

PRISM Collection at UCF Libraries

○ Political & Rights Issues & Social Movements (PRISM) collection: Part of the PALMM (Publication of Archival Library & Museum Materials).

	 - PDF Document (10 M)
Title	The meaning of the Soviet trials
Creator	Yaroslavskii, Emelian 1878-1943 Bukharin, Nikolai Ivanovich 1888-1938 defendant Foster, William Z. 1881-1961 Yagoda, Genrikh Grigorovich 1891-1938 defendant Rykov, Aleksei Ivanovich 1881-1938 defendant
Subject(s)	Bukharin, Nikolai Ivanovich 1888-1938 -- Trials, litigation, etc Trials (Political crimes and offenses) -- Russia (Federation) -- Moscow
Description	46, [1] p. ; 19 cm.
Publisher	New York : Workers Library Publishers
Date	1938
Type	Book
Source	http://ucf.catalog.fcla.edu/permalink.jsp?29CF000388014
Language	English
Rights	All rights to images are held by the respective holding institution. This image is posted publicly for non-profit educational uses, excluding printed publication. For permission to reproduce images and/or for copyright information contact Special Collections and University Archives, University of Central Florida Libraries, (407) 823-2576. http://library.ucf.edu/SpecialCollections/
Physical Location	UCF Libraries Special Collections - 5th Floor -- HX15.V35 no.829
Identifier	DP0000635.pdf
PID	671335
Number of views	29
Related collections	All Collections > PRISM - Political & Rights Issues & Social Movement Items by Contributor > University of Central Florida Libraries Collections by Media > Text > PRISM - Political & Rights Issues & Social Movement Collections by Type > Books > PRISM - Political & Rights Issues & Social Movement Collections by Topic > History > PRISM - Political & Rights Issues & Social Movement

A PRISM record in DigiTool: <http://purl.flvc.org/FCLA/DT/671335>

The DigiTool DC File

```
<?xml version="1.0" encoding="UTF-8"?>
<record xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>The meaning of the Soviet trials</dc:title>
  <dcterms:alternative/>
  <dcterms:alternative/>
  <dcterms:alternative/>
  <dc:creator>Yaroslavskii, Emel'ian 1878-1943 </dc:creator>
  <dc:creator>Bukharin, Nikolai Ivanovich 1888-1938 defendant</dc:creator>
  <dc:creator>Foster, William Z. 1881-1961</dc:creator>
  <dc:creator>Yagoda, Genrikh Grigorévich 1891-1938 defendant</dc:creator>
  <dc:creator>Rykov, Aleksei Ivanovich 1881-1938 defendant</dc:creator>
  <dc:subject>Bukharin, Nikolai Ivanovich 1888-1938 -- Trials, litigation,
  etc</dc:subject>
  <dc:subject>Trials (Political crimes and offenses) -- Russia (Federation) --
  Moscow</dc:subject>
  <dc:subject/>
  <dc:subject/>
  <dc:subject/>
  <dc:subject/>
  <dc:subject/>
  <dc:subject/>
  <dc:description>46, [1] p. ; 19 cm.</dc:description>
  <dc:description/>
  <dc:publisher>New York : Workers Library Publishers</dc:publisher>
  <dc:contributor/>
  <dc:contributor/>
  <dc:contributor/>
  <dc:contributor/>
  <dc:contributor/>
  <dc:contributor/>
  <dc:date>1938</dc:date>
  <dc:type>Book</dc:type>
  <dc:identifier>DP0000635.pdf</dc:identifier>
  <dc:source>http://ucf.catalog.fcla.edu/permalink.jsp?29CF000388014</dc:source>
  <dc:language>English</dc:language>
  <dc:language/>
  <dc:rights>All rights to images are held by the respective holding institution. This
  image is posted publicly for non-profit educational uses, excluding printed
  publication. For permission to reproduce images and/or for copyright information
  contact Special Collections and University Archives, University of Central Florida
  Libraries, (407) 823-2576. http://library.ucf.edu/SpecialCollections/</dc:rights>
  <dcterms:physicalLocation>UCF Libraries Special Collections - 5th Floor -- HX15.V35
  no.829</dcterms:physicalLocation>
</record>
```

dc:title (1. If subtitle available, will be in the same field)

dc:creator (5. Their relationship unknown, name year and role undifferentiated)

dc:subjects (2. Subdivisions not available. 5 empty subject fields can be disregarded)

dc:description (physical description is not distinguished from general description)

dc:publisher (place and publisher mixed together)

dc:date (unclear what type of date)

dcterms:physicalLocation (impossible to distinguish location and call number)

Wrapped in <xb><mds><md><value> in the DigiTool generated file

Converted MODS (initial test)

```
<?xml version="1.0" encoding="UTF-8"?>
- <mods version="3.4" xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-4.xsd" xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:flvc="info:flvc/manifest/v1" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.loc.gov/mods/v3">
  - <titleInfo>
    <title>The meaning of the Soviet trials</title>
  </titleInfo>
  <identifier type="digitool">671335</identifier>
  - <name>
    <namePart>Yaroslavskii, Emel'ian 1878-1943</namePart>
    - <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
  - <name>
    <namePart>Bukharin, Nikolai Ivanovich 1888-1938 defendant</namePart>
    - <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
  - <name>
    <namePart>Foster, William Z. 1881-1961</namePart>
    - <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
  - <name>
    <namePart>Yagoda, Genrikh Grigorévich 1891-1938 defendant</namePart>
    - <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
  - <name>
    <namePart>Rykov, Aleksei Ivanovich 1881-1938 defendant</namePart>
    - <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
  <genre>book</genre>
```

Issues:

Subtitle is not marked up (when there is a subtitle);

Non-filling characters not marked up;

Name type is not available;

Relationship among names unclear;

Author/name year is not marked up;

Author/name role is not marked up;

When there are more than one creator, it creates problem in MARC display (multiple 1XXs; 1XX/7XX unclear).

Converted MODS (initial test, cont.)

```
- <originInfo>
  <publisher>New York : Workers Library Publishers</publisher>
  <dateIssued keyDate="yes" encoding="w3cdtf">1938</dateIssued>
</originInfo>
- <language>
  <languageTerm type="text">English</languageTerm>
</language>
<note>46, [1] p. ; 19 cm.</note>
- <subject>
  <topic>Bukharin, Nikolai Ivanovich 1888-1938 -- Trials, litigation, etc</topic>
</subject>
- <subject>
  <topic>Trials (Political crimes and offenses) -- Russia (Federation) -- Moscow</topic>
</subject>
- <relatedItem type="original">
  - <location>
    <url>http://ucf.catalog.fcla.edu/permalink.jsp?29CF000388014</url>
    </location>
    <identifier type="uri">http://ucf.catalog.fcla.edu/permalink.jsp?29CF000388014</i
  </relatedItem>
- <location>
  <physicalLocation>UCF Libraries Special Collections - 5th Floor -- HX15.V35 no.829</physicalLocation>
</location>
- <location displayLabel="purl">
  <url>http://purl.flvc.org/FCLA/DT/671335</url>
</location>
<accessCondition>All rights to images are held by the respective holding institution. This image is posted publicly for
non-profit educational uses, excluding printed publication. For permission to reproduce images and/or for
copyright information contact Special Collections and University Archives, University of Central Florida
Libraries, (407) 823-2576. http://library.ucf.edu/SpecialCollections/</accessCondition>
<typeOfResource>still image</typeOfResource>
</mods>
```

Issues:

Place and Publisher cannot be distinguished in <publisher>;

Physical description is under <note>;

Subject subdivisions not available;

Links may need some explanation (for the public);

PhysicalLocation has mixed information.

Converted MARC (initial test, Cont.)

○ Issues (in the initial test):

○ Information not mapped to the most accurate field:

- The main author/creator is mapped to 720, which should go to 100. (LC DC-MODS-MARC stylesheets map all names to 720 to accommodate all types of names, e.g., controlled and uncontrolled, person/corporate/conference)
- Additional authors/creators and contributors are mapped to 720, which should go to 700. (Impossible for LC DC-MODS stylesheet to distinguish between main entries and added entries)
- Physical descriptions mapped to note (500 etc.), but not 300. (Physical and other descriptions are not distinguished in DC.)

○ Indicator problems:

- Non filling characters need to be identified. (LC DC-MODS-MARC stylesheets do not take care of non-filling characters)
- First indicator for 245, 100, 700 need to be corrected. (LC DC-MODS-MARC stylesheets do not deal with the relationship between title and author, main entry and added entries)

○ Subfield problems:

- The publisher place should be mapped to 260 |a, not |b. (LC DC-MODS-MARC stylesheets do not distinguish publication place from the publisher)
- For all the creators/authors/names, the year range need to be under |d, role needs to be under |c.
- The stylesheets are unable to distinguish the subtitle (245 |b) from title (245 |a).

○ Inadequate information:

- The collection name needs to be preserved (if possible) in 830 0.
- The stylesheets are unable to produce a 245 |c. Even we may bypass 245 |c due to the inadequate dc records info, we need to correct 245 |b.

Options to Improve Metadata: XSLT customization

- Can have customized XSLT stylesheets, but it will require an environment to test and run the code on the data source, and preferably, a programmer.
- **DC-MODS and MODS-MARC stylesheets Adjustments (Some possible ideas)**
 - Adjust the **main entry and added entry** (1XX, 7XX) logic. Map the first name to 1XX (**Element Positioning**), other names to 7XX. However how do you know if it's a person (100/700), an organization (110/710) or a conference (111/711)?
 - Apply data patterns and automatically distinguish and mark up personal, corporate and conference **name types** (may cover the majority of the situations, but impossible to make all records right. Need review);
 - Generate mark-up code for **name year** and **role** based on data patterns for year and role terms;
 - Generate mark-up code for **subtitles** based on whether a ":" exists in the title;
 - Generate mark-up code for **publication place** based on whether a ":" exists in the publisher field;

Options to Improve Metadata: XSLT customization

(Cont.)

- To identify **field relationships** and have the right indicators for 100 and 245 based on "if an author exists."
 - Other adjustments: mark up **non-filling characters**...
 - **Need comprehensive review of the results!**
 - Some of these adjustments cannot produce one hundred percent correct results (because not all data can fall into a pattern); it might not be realistic to apply them to a national or consortia level stylesheet. They seem to be more feasible on a library collection level. However, what do you think?
-
- **MARC-MODS, MODS-MARC stylesheets**
 - MARC-MODS conversion produces much better results (e.g. for the Florida Heritage collection in DigiTool). However, a question on local subjects was raised because MARC-MODS XSLT does not take local subjects (690).
 - **Subject mapping change (by FLVC)**
 - **MARC-MODS:** Mapped local **690** to **<subject><topic>**
 - Added subject authority "sears": **6XX _8** mapped to **<subject authority="sears">**
 - **MODS-MARC:** **<subject authority="sears">** mapped to **6XX _8**

Options to Improve Metadata: MODS Editing

- Post-transformation MODS Editing
- FLVC recommended Notepad++ for MODS editing

- **Notepad++**
 - Free text and source code editor for Windows
 - Can be downloaded at: <http://notepad-plus-plus.org/>
 - Download Executable zip file and install

 - **Some characteristics:**
 - Set different languages (XML, HTML, CSS, PHP, Java...)
 - Syntax color coding
 - Tree view of the files and folders
 - Enable tag autocompletion (in set preference)
 - Can open, save and close multiple files at one time (save all, close all...)
 - Plug in: download plug in, such as ftp, XML tools
 - Editing: Find, replacement, can use regular expressions, macros...
 - Check xml for well-formedness, validate against xml schema

- PRISM collection: Edited and marked up 847 MODS XML records using Notepad++ which were generated from Dublin Core (DC).

MODS Editing: Distinguish between the Main Entry and Added Entries

- When there are more than one creator (`<name><role><roleTerm>` with `type="text"` value "creator," or, `<name><role><roleTerm>` with `type="code"` value "cre"), there can be multiple 100s. However 100 is not repeatable.
- Only keep the first creator marked up with "creator" or "cre" in MODS file, and change the other name types to "contributor." Or just leave the other names' types blank (depending on the logic in the stylesheet).

The screenshot shows a Notepad++ window with an XML file open. The XML code is as follows:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-1
3 <titleInfo>
4 <title>Marxism vs. liberalism, an interview [between] Joseph Stalin [and] H. G. We
5 </titleInfo>
6 <identifier type="digitool">
7 2683749</identifier>
8 <name>
9 <namePart>Stalin, Joseph</namePart><namePart type="date">1878-1953</namePart>
10 <role>
11 <roleTerm type="text" authority="marcrelator">creator</roleTerm>
12 </role>
13 </name>
14 <name>
15 <namePart>Wells, H. G. (Herbert George)</namePart><namePart type="date">1866-1946<
16 <role>
17 <roleTerm type="text" authority="marcrelator">creator</roleTerm>
```

The Find dialog box is open with the following settings:

- Find what: `</roleTerm>`
- Find Next
- Count
- Find All in All Opened Documents
- Find All in Current Document
- Close
- Match whole word only:
- Match case:
- Wrap around:
- Search Mode: Normal
- Direction: Up, Down
- Transparency: On losing focus, Always

The search results window shows the following hits:

```
Find result - 976 hits
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT2683749.xml (2 hits)
Line 11: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
Line 17: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT2683751.xml (4 hits)
Line 11: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
Line 17: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
Line 23: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
Line 29: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT2683753.xml (1 hit)
Line 11: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT2683755.xml (1 hit)
Line 11: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT2683757.xml (2 hits)
Line 14: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
Line 20: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT2683759.xml (1 hit)
Line 11: <roleTerm type="text" authority="marcrelator">creator</roleTerm>
```

Find what: `</roleTerm>`

When there are more than "1 hit" (1 creator) in a record, keep the first creator "creator," change the others to "contributor" or leave them un-marked-up.

MODS Editing: Identify Pattern and Mark Up Data (Year)

- Year pattern: yyyy-yyyy; yyyy- yyyy; yyyy-; yyyy; b. yyyy

(Find pattern by searching `<namePart>` in Notepad++ and sort the results in a spreadsheet)

- Mark-up editing, for example:

```
<name>  
  <namePart>De Leon, Daniel 1852-1914</namePart>  
  ...  
</name>
```

Replace with:

```
<name>  
  <namePart>De Leon, Daniel</namePart>  
  <namePart type="date">1852-1914</namePart>  
  ...  
</name>
```

MODS Editing: Identify Pattern and Mark Up Data (Year, Cont.)

- Use Regular Expression in Notepad++ to perform find/replacement function and mark up the year;
- () is used to tag a match. Tagged matches can be referred to as \1, \2 etc.

Find What: `(([0-9]+)-([0-9]+))</namePart>`

Replace with: `</namePart><namePart type="date">\1-\2</namePart>`

CFDT212128.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:flvc="info:flvc/manifest/v1"
  <titleInfo>
    <title?What means this strike? address delivered by Daniel De Leon in the City Hall of New Bedford, Massachusetts, February 11, 1898.</title>
  </titleInfo>
  <identifier type="digitool">
    212128</identifier>
  <name>
    <namePart>De Leon, Daniel</namePart><namePart type="date">1852-1914</namePart>
    <roleTerm type="text" authority="marcrelator">creator</roleTerm>
  </role>
</name>
<genre>book</genre>
```

Find result - 264 hits

Search " (([0-9]+)-([0-9]+))</namePart>" (264 hits in 248 files)

- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT212128.xml (1 hit)
Line 9: <namePart>De Leon, Daniel 1852-1914</namePart>
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT212130.xml (1 hit)
Line 9: <namePart>Simons, A. M. (Algie Martin) 1870-1950</namePart>
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT246063.xml (1 hit)
Line 12: <namePart>De Leon, Daniel 1852-1914</namePart>
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT334777.xml (1 hit)
Line 9: <namePart>Haywood, Big Bill 1869-1928</namePart>

Replace

Find what: `(([0-9]+)-([0-9]+))</namePart>`

Replace with: `</namePart><namePart type="date">\1-\2</namePart>`

Match case Wrap around

Search Mode: Regular expression

Direction: Down

Replace All in All Opened Documents

Replace in Opened Files: 264 occurrences replaced.

MODS Editing: Identify Pattern and Mark Up Data (Year, Cont.)

○ `<namePart>Vail, Charles H. (Charles Henry) b. 1866</namePart>`

Replaced With:

`<namePart>Vail, Charles H. (Charles Henry)</namePart><namePart type="date">b. 1866</namePart>`

The screenshot shows a Notepad++ window editing an XML file. The XML content is as follows:

```
<?xml version="1.0" encoding="UTF-8" ?>
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xdm="http://www.loc.gov/mods/v3#xdm">
  <titleInfo>
    <title>Principles of scientific socialism </title>
  </titleInfo>
  <identifier type="digitool">
    212124</identifier>
  <name>
    <namePart>Vail, Charles H. (Charles Henry)</namePart><namePart type="date">b. 1866</namePart>
    <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
  <genre>book</genre>
</mods>
```

A purple oval highlights the line containing the name and date in the XML. A blue callout box on the right contains the following text:

Find What: `b. ([0-9]+)</namePart>`
Replace with: `</namePart><namePart type="date">b. \1</namePart>`

The Replace dialog box is open, showing the search and replace patterns. The 'Find what' field contains `b. ([0-9]+)</namePart>` and the 'Replace with' field contains `</namePart><namePart type="date">b. \1</namePart>`. The 'Regular expression' radio button is selected. The 'Replace All' button is highlighted. The status bar at the bottom indicates 'Replace in Opened Files: 4 occurrences replaced.'

Below the XML editor, a 'Find result - 4 hits' pane shows the following search results:

- Search " b. ([0-9]+)</namePart>" (4 hits in 4 files)
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT212124.xml (1 hit)
Line 9: <namePart>Vail, Charles H. (Charles Henry) b. 1866</namePart>
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT358321.xml (1 hit)
Line 15: <namePart>Levy, Oscar b. 1867</namePart>
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT358379.xml (1 hit)
Line 9: <namePart>Burns, Emile b. 1889</namePart>
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT1927447.xml (1 hit)
Line 9: <namePart>Trojanovski, A. (Aleksandr) b. 1882</namePart>
- Search "1866</namePart>" (1 hit in 1 file)
- Search " b. ([0-9]+)</namePart>" (4 hits in 4 files)
- Search "([0-9]+)</namePart>" (273 hits in 254 files)
- Search "([0-9]+)</namePart>" (204 hits in 198 files)

MODS Editing: Identify Pattern and Mark Up Data (Role Terms)

- Check pattern (click "Find All in All Opened Documents")

Find What: </namePart>

The screenshot shows the Notepad++ interface with an XML file open. The XML content is as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:flvc="info:flvc/manifest/v1">
  <titleInfo>
    <title>Industrial socialism</title>
  </titleInfo>
  <identifier type="digitool">
    334777</identifier>
  <name>
    <namePart>Haywood, Big Bill</namePart><namePart type="date">1869-1928</namePart>
    <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
  <name>
    <namePart>Bohn, Frank 1878- joint author</namePart>
    <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
</mods>
```

The Find dialog box is open, showing the search pattern </namePart> and the "Find All in All Opened Documents" button circled in purple. The search results at the bottom show 1591 hits across multiple files, with the first few lines of the XML content highlighted in green.

MODS Editing: Identify Pattern and Mark Up Data (Role Terms, Cont.)

- Copy the result to a spreadsheet

A	B	C	D	E	F	G	H	I	J	K	L
C:\My Documents\sl	landora\PrismEdit\ucf_smc_mods\CFDT212124.xml (2 hits)										
	Line 9:	<namePart>Vail, Charles H. (Charles Henry)</namePart><namePart type="date">b. 1866</namePart>									
	Line 9:	<namePart>Vail, Charles H. (Charles Henry)</namePart><namePart type="date">b. 1866</namePart>									
C:\My Documents\sl	landora\PrismEdit\ucf_smc_mods\CFDT212126.xml (2 hits)										
	Line 9:	<namePart>McClure, William Scholl</namePart>									
	Line 15:	<namePart>Albany Press Club</namePart>									
C:\My Documents\sl	landora\PrismEdit\ucf_smc_mods\CFDT212128.xml (2 hits)										
	Line 9:	<namePart>De Leon, Daniel</namePart><namePart type="date">1852-1914</namePart>									
	Line 9:	<namePart>De Leon, Daniel</namePart><namePart type="date">1852-1914</namePart>									
C:\My Documents\sl	landora\PrismEdit\ucf_smc_mods\CFDT212130.xml (2 hits)										
	Line 9:	<namePart>Simons, A. M. (Algie Martin)</namePart><namePart type="date">1870-1950</namePart>									
	Line 9:	<namePart>Simons, A. M. (Algie Martin)</namePart><namePart type="date">1870-1950</namePart>									

- Sort the value in the second column to see patterns

Line 15:	<namePart>Raushenbush, Carl - editor</namePart>
Line 15:	<namePart>Real, James joint author</namePart>
Line 15:	<namePart>Red International of Labor Unions British Bureau</namePart>
Line 15:	<namePart>Russian S.F.S.R Laws, etc</namePart>
Line 15:	<namePart>Seidman, Joel Isaac 1906 joint author</namePart>
Line 15:	<namePart>Shachtman, Max 1903- tr</namePart>
Line 15:	<namePart>Shachtman, Max 1903- tr</namePart>
Line 15:	<namePart>Shiel, M. P. (Matthew Phipps) 1865-1947 translator</namePart>
Line 15:	<namePart>Shils, Edward</namePart><namePart type="date">1910-1995</namePart>
Line 15:	<namePart>Shils, Edward</namePart><namePart type="date">1910-1995</namePart>
Line 15:	<namePart>Shore, Merle</namePart>
Line 15:	<namePart>Simons, A. M. - translator</namePart>
Line 15:	<namePart>Singer, J. David (Joel David) 1925- Surprise attack</namePart>
Line 15:	<namePart>Socialist labor party</namePart>
Line 15:	<namePart>Socialist Labor Party</namePart>

MODS Editing: Identify Pattern and Mark Up Data (Role Terms, Cont.)

○ Author Role Terms Pattern

Role terms	Role terms (after being controlled)
defendant	defendant
illus.	illustrator
- illus	
joint author	author
ed	editor
joint ed	
editor	
- editor	
translator	translator
- translator	
- translator	
tr (at the end)	

- Clean up and standardize the terms (find/replace);

- Edit the mark up, for example:

Replace:

```
<name>  
  <namePart>McMillan, Hugh translator</namePart>  
</name>
```

With:

```
<name>  
  <namePart>McMillan, Hugh </namePart><role><roleTerm type="text"  
  authority="marcrelator">translator</roleTerm></role>  
</name>
```

MODS Editing: Identify Pattern and Mark Up Data (Role Terms, Cont.)

○ In Notepad++

Find what: translator</namePart>

Replace with: </namePart><role><roleTerm type="text"
authority="marcrelator">translator</roleTerm></role>

The screenshot shows the Notepad++ interface with an XML document open. The XML content is as follows:

```
10 | 358317</identifier>
11 | <name>
12 | <namePart>Gorter, Herman</namePart><namePart type="date">1864-1927</namePart>
13 | <role>
14 | <roleTerm type="text" authority="marcrelator">creator</roleTerm>
15 | </role>
16 | </name>
17 | <name>
18 | <namePart>McMillan, Hugh</namePart><role><roleTerm type="text" authority="marcrelator">translator</roleTerm></role>
19 | </name>
20 | <genre>book</genre>
21 | <originInfo>
22 | <publisher>Glasgow : Socialist Information and Research Bureau
23 | <dateIssued encoding="w3cdtf" keyDate="yes">1920</dateIssued>
24 | </originInfo>
25 | <language>
26 | <languageTerm type="text">English</languageTerm>
```

A purple oval highlights the text on line 18: `<namePart>McMillan, Hugh</namePart><role><roleTerm type="text" authority="marcrelator">translator</roleTerm></role>`. A Replace dialog box is open in the foreground, with the following settings:

- Find what: `translator</namePart>`
- Replace with: `</namePart><role><roleTerm type="text" authority="marcrelator">translator</roleTerm></role>`
- Match whole word only:
- Match case:
- Wrap around:
- Search Mode: Normal
- Direction: Down
- Transparency: On losing focus

The dialog box also shows "Replace in Opened Files: 20 occurrences replaced." and "Replace All in All Opened Documents" button.

Below the XML editor, a "Find result - 20 hits" window shows the search results for "translator</namePart>" across 17 files, with the relevant line highlighted in each file.

Please notice that: when there is a role term after "yyyy-yyyy," the year could not be marked up in the previous "year" replacement. Can replace it after the role term is marked up.

MODS Editing: Identify Pattern and Mark Up Data (Names)

- Distinguished between personal names, corporate names and conferences;
 - Find: <namePart>, copy the result to a spreadsheet, sort data and identify some patterns;
 - What's in spreadsheet:

	B	C
Line 12:	<namePart>Brody, Alter</namePart><namePart type="date">1895-</namePart>	
Line 12:	<namePart>Bukharin, Nikolai Ivanovich</namePart><namePart type="date">1888-1938</namePart>	
Line 12:	<namePart>Cole, G. D. H. (George Douglas Howard)</namePart><namePart type="date">1889-1959</namePart>	
Line 12:	<namePart>Confederación de Trabajadores de México</namePart>	
Line 12:	<namePart>De Leon, Daniel</namePart>	
Line 12:	<namePart>De Leon, Daniel</namePart><namePart type="date">1852-1914</namePart>	
Line 12:	<namePart>Drob, Judah</namePart>	
Line 12:	<namePart>Fourth International Conference (1946 : Brussels)</namePart>	
Line 12:	<namePart>Gorter, Herman</namePart><namePart type="date">1864-1927</namePart>	
Line 12:	<namePart>Greater New York Committee for a Sane Nuclear Policy</namePart>	

MODS Editing: Identify Pattern and Mark Up Data (Names, Cont.)

Name Pattern (if <namePart> contains)	Add Code
Party, Partei, partiia, partiï	<name type="corporate">
Committee	
International	
Council	
National	
agency	
Service	
School	
World	
Association	
society	
department	
Mission	
Union	
League	
bund	
scientists of	
Conference	<name type="conference">
Congress	
Others	<name type="personal">

Not comprehensive. Need to check the results!!

MODS Editing: Identify Pattern and Mark Up Data (Names, Cont.)

- Find all <namePart> elements which contain "Committee"

Find what: <name>\n <namePart>.*Committee.*</namePart>

The screenshot shows a Notepad++ window with an XML document open. The XML content is as follows:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:flvc="info:flvc/manifest/v1
3 <titleInfo>
4   <title>15 years of the Communist International</title>
5 </titleInfo>
6 <identifier type="digitool">
7   363432</identifier>
8 <name>
9   <namePart>Communist International Executive Committee</namePart>
10  <role>
11    <roleTerm type="text" authority="marcrelator">creator</roleTerm>
12  </role>
13 </name>
14 <genre>book</genre>
15 <originInfo>
16   <publisher>New York: Workers Library Publishers</publisher>
17   <dateIssued encoding="w3cdtf" keyDate="yes">1934</dateIssued>
```

The Find dialog box is open, showing the search pattern: <name>\n <namePart>.*Committee.*</namePart>. The "Regular expression" search mode is selected. The "Find Next" button is highlighted. The search results pane at the bottom shows 2 hits in 2 files, with the first hit at Line 8 of CFDT363432.xml.

MODS Editing: Identify Pattern and Mark Up Data (Names, Cont.)

Find what: `<name>\n <namePart>(.*Committee.*)</namePart>`

Replace with: `<name type="corporate">\n <namePart>\1</namePart>`

(Note: will only work when the corporate name is the first name. May need to manually add: `type="corporate"` for `<name>`.)

The screenshot shows a Notepad++ window with an XML document open. The XML content is as follows:

```
1 <?xml version="1.0" encoding="UTF-8" ?>
2 <mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:flvc="info:flvc/manifest/v1"
3 <titleInfo>
4 <title>15 years of the Communist International</title>
5 </titleInfo>
6 <identifier type="digitool">
7 363432</identifier>
8 <name type="corporate">
9 <namePart>Communist International Executive Committee</namePart>
10 <role>
11 <roleTerm type="text" authority="marcrelator">creator</roleTerm>
12 </role>
13 </name>
14 <genre>book</genre>
15 <originInfo>
16 <publisher>New York: Workers Library Publishers</publisher>
17 <dateIssued encoding="w3cdtf" keyDate="yes">1934</dateIssued>
```

The XML editor shows a tree view on the left. The `<name type="corporate">` tag on line 8 and the `<namePart>Communist International Executive Committee</namePart>` tag on line 9 are circled in purple. A "Replace" dialog box is open in the foreground, also with purple circles around the search and replace text. The search text is `<name>\n <namePart>(.*Committee.*)</namePart>` and the replace text is `<name type="corporate">\n <namePart>\1</namePart>`. The dialog box has tabs for "Find", "Replace", "Find in Files", and "Mark". The "Replace" tab is active. The "Find what" and "Replace with" fields are circled in purple. The "Replace" dialog box also has buttons for "Find Next", "Replace", "Replace All", "Replace All in All Opened Documents", and "Close". There are checkboxes for "Match whole word only", "Match case", and "Wrap around". The "Wrap around" checkbox is checked. There are also sections for "Search Mode" (Normal, Extended, Regular expression), "Direction" (Up, Down), and "Transparency" (On losing focus, Always). The "Regular expression" radio button is selected, and the "matches newline" checkbox is checked. The "Replace in Opened Files: 2 occurrences replaced." message is visible at the bottom of the dialog box.

MODS Editing: Identify Pattern and Mark Up Data (Names, Cont.)

- After checking corporate and conference names, replace all remaining names:

<name>
<namePart>

With

<name type="personal">
<namePart>

**Need to review all the names!!
Make edits when needed.**

The screenshot shows a Notepad++ window editing an XML file. The XML content is as follows:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:flvc="info:flvc/manifest/v1
3 <titleInfo>
4 <title>Socialism and superior brains : a reply to Mr. Mallock</title>
5 </titleInfo>
6 <identifier type="digitool">
7 326838</identifier>
8 <name type="personal">
9 <namePart>Shaw, Bernard</namePart>
10 <role>
11 <roleTerm type="text" authority="marcrelator">creator</roleTerm>
12 </role>
13 </name>
14 <genre>book</genre>
15 <originInfo>
16 <publisher>London : Fabian Society</publisher>
17 <dateIssued encoding="w3cdtf" keyDate="yes">1909</dateIssued>
```

The Replace dialog box is open, showing the following configuration:

- Find what: <name> <namePart>
- Replace with: <name type="personal"> <namePart>
- Match case: checked
- Wrap around: checked
- Search Mode: Extended (n, v, t, l, |, \, ...)
- Direction: Down
- Transparency: On losing focus

The status bar at the bottom of the dialog indicates: "Replace in Opened Files: 1107 occurrences replaced."

MODS Editing: Identify Pattern and Mark Up Data (Subtitle)

- Find all titles with subtitles (find “:” in title field).

Find what: <title>(.*:.)</title>

The screenshot shows a Notepad++ window with a MODS XML file open. The XML content is as follows:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:flvc="info:flvc/manifest/v1"
3 <titleInfo>
4 <title>Socialism : a paper read before the Albany Press Club "Socialist Night" <
5 </titleInfo>
6 <identifier type="digitool">
7 212126</identifier>
```

A search dialog box is open over the XML, with the search pattern `<title>(.*:.)</title>` entered in the "Find what:" field. The search results pane at the bottom shows 225 hits across 223 files. The first few results are:

- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT212126.xml (1 hit)
Line 4: <title>Socialism : a paper read before the Albany Press Club "S
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT247096.xml (1 hit)
Line 4: <title>Unity : an address delivered by Daniel De Leon at New Pyt
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT326838.xml (1 hit)
Line 4: <title>Socialism and superior brains : a reply to Mr. Mallock</
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT331852.xml (1 hit)
Line 4: <title>The trial of Ferrer : a clerical-judicial murder</title>
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT331854.xml (1 hit)
Line 4: <title>Robert Owen : social reformer</title>
- C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\CFDT331858.xml (2 hits)
Line 4: <title>N. Lenin : his life and work</title>
Line 7: <title>Nicolai Lenin : his life and work</title>

It returns 225 hits.

MODS Editing: Identify Pattern and Mark Up Data (Subtitle, Cont.)

○ For example:

Replace

<titleInfo>

<title>Socialism : a paper read before the Albany Press Club "Socialist Night" </title>

</titleInfo>

With:

<titleInfo>

<title>Socialism : <subtitle> a paper read before the Albany Press Club "Socialist Night"</subtitle>

</title>

</titleInfo>

May leave out ":" in title.

Find what: <title>(.*):(.*)</title>

Replace with:

<title>\1:<subtitle>\2</subtitle></title>

The screenshot shows a Notepad++ window with an XML file open. The XML content is as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:flvc="info:flvc/manifest/v1">
  <titleInfo>
    <title>Socialism :<subtitle> a paper read before the Albany Press Club "Socialist Night"</subtitle></title>
  </titleInfo>
  <identifier type="digitool">

```

The line containing the title and subtitle is circled in purple. Below the XML editor, a search results pane shows 225 hits for the search pattern "<title>(.*):(.*)</title>".

Overlaid on the bottom right is a "Replace" dialog box. The "Find what" field contains "<title>(.*):(.*)</title>" and the "Replace with" field contains "<title>\1:<subtitle>\2</subtitle></title>". Both fields are circled in purple. The "Regular expression" radio button is selected. The "Replace All" button is highlighted in blue.

MODS Editing: Identify Pattern and Mark Up Data (non-filling characters)

○ For example:

- ... Is this a war for freedom? `<nonSort>...</nonSort>`
- A key to survival `<nonSort>A </nonSort>`
- An alternative to war `<nonSort>An </nonSort>`
- The theory of the Cuban Revolution `<nonSort>The </nonSort>`

Find what: `<title>The`
Replace with:
`<title><nonSort>The </nonSort>`

The screenshot shows a Notepad++ window with a MODS XML file open. The XML content is as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:flvc="info:flvc/manifest/v1">
  <titleInfo>
    <title><nonSort>The </nonSort> trial of Ferrer a clerical-judicial murder</title>
  </titleInfo>
  <identifier type="digitool">

```

The search and replace dialog box is open, showing the following settings:

- Find what: `<title>The`
- Replace with: `<title><nonSort>The </nonSort>`
- Match whole word only:
- Match case:
- Wrap around:
- Search Mode: Normal
- Direction: Down
- Transparency: On losing focus

The search results pane at the bottom shows 258 hits across 257 files. The results are as follows:

File Path	Line	Content
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\	Line 4:	<code><title>The trial of Ferrer a clerical-judicial murder</title></code>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\	Line 4:	<code><title>The ABC of socialism</title></code>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\	Line 4:	<code><title>The Catholic church and society</title></code>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\	Line 4:	<code><title>The American socialists and their times</title></code>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\	Line 4:	<code><title>The student's epitome of socialism</title></code>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\	Line 4:	<code><title>The bankruptcy of the American Republic</title></code>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods\	Line 4:	<code><title>The Revolutionary act; military and political</title></code>

MODS Editing: Identify Pattern and Mark Up Data (Publication Place)

○ Identify publication place pattern

The screenshot shows a Notepad++ window with an XML document. The XML content includes a `<publisher>` tag with the value "New York : New York Labor News". A purple oval highlights the text "New York : New York Labor News" in the XML. A purple callout box above the XML contains the text "Find what: <publisher>(.*):(.*)</publisher>". Below the XML, a "Find" dialog box is open, with the same search pattern entered in the "Find what:" field, also circled in purple. The dialog box includes options for "Match whole word only", "Match case", "Wrap around", "Search Mode" (Normal, Extended, Regular expression), "Direction" (Up, Down), and "Transparency" (On losing focus, Always).

○ Publisher place pattern, for example:

"New York

[New York

[New York]

New York City

New York, [N.Y.]

New York, N.Y.

New York, NY

(all need to change to New York, N. Y.)

MODS Editing: Identify Pattern and Mark Up Data (Publication Place)

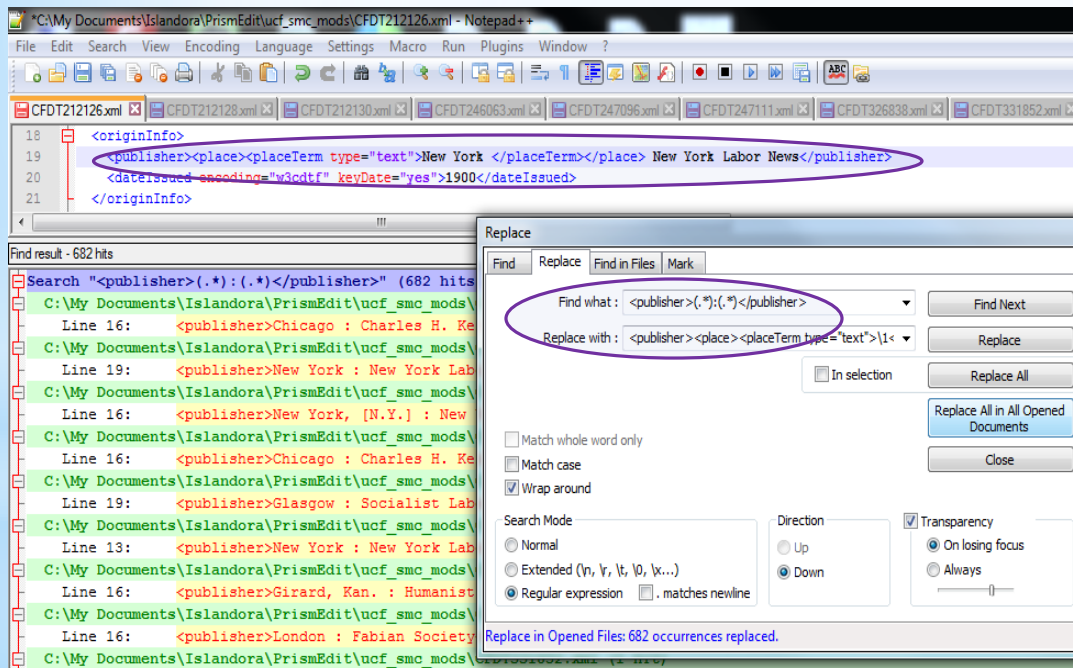
○ For example:

Replace

<publisher>New York : New York Labor News</publisher>

with

<publisher><place><placeTerm type="text">New York </placeTerm></place> New York Labor News</publisher>



Find what:

<publisher>(.*):(.*)</publisher>

Replace with:

<publisher><place><placeTerm type="text">\1</placeTerm></place>\2</publisher>

Need to standardize and clean the place terms before the mark-up replacement, e.g. change "New York" to "New York, N.Y."

MODS Editing: Identify Pattern and Mark Up Data (Subject)

- Mark up topical, temporal, geographic, genre subdivisions for subjects, and personal/corporate/conference name subjects
- Subdivision patterns: Use test set to find pattern (don't mess with the "edit-in-progress" set)

Find what: `<topic>(.*)--(.*?)</topic>`
Replace with
`<topic>--\2</topic>`

The screenshot shows a Notepad++ window with an XML document. The XML content includes several `<subject>` elements with `<topic>` subdivisions. A purple circle highlights a specific `<subject>` element. A 'Replace' dialog box is open, showing the search pattern `<topic>(.*)--(.*?)</topic>` and the replacement `<topic>--\2</topic>`. The 'Find' tab is active, and the 'Replace All' button is highlighted. Below the dialog, a 'Find result' window shows 859 hits in 540 files, with a list of search results.

```
Find result - 859 hits
Search "<topic>(.*?)--(.*?)</topic>" (859 hits in 540 files)
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods_test2\CFDT212130.xml
Line 33: <topic>Textile workers -- Massachusetts</topic>
C:\My Documents\Islandora\PrismEdit\ucf_smc_mods_test2\CFDT212130.xml
Line 24: <topic>Labor -- United States</topic>
Line 27: <topic>Socialism -- United States</topic>
Line 30: <topic>United States -- Economic conditions</topic>
Line 33: <topic>United States -- Social conditions -- 1865-1918</topic>
Line 36: <topic>Working class -- United States</topic>
```

Then find "`<topic>--(.*?)</topic>`" to get subdivisions only, copy to a spreadsheet to identify patterns.

MODS Editing: Identify Pattern and Mark Up Data (Subject, cont.)

○ Subject subdivision patterns

Line 33:	<topic>-- Massachusetts</topic>
Line 24:	<topic>-- United States</topic>
Line 27:	<topic>-- United States</topic>
Line 30:	<topic>-- Economic conditions</topic>
Line 33:	<topic>-- 1865-1918</topic>
Line 36:	<topic>-- United States</topic>
Line 24:	<topic>-- Outlines, syllabi, etc</topic>
Line 27:	<topic>-- Outlines, syllabi, etc</topic>
Line 30:	<topic>-- Philosophy</topic>
Line 24:	<topic>-- Revolution, 1909</topic>
Line 24:	<topic>-- Great Britain</topic>
Line 30:	<topic>-- Great Britain</topic>
Line 30:	<topic>-- Christianity</topic>
Line 24:	<topic>-- Economic aspects</topic>
Line 30:	<topic>-- Biography</topic>
Line 33:	<topic>-- Revolution, 1917-1921</topic>
Line 36:	<topic>-- Biography</topic>
Line 27:	<topic>--Biography </topic>
Line 30:	<topic>-- Catholic Church</topic>
Line 36:	<topic>-- Controversial literature</topic>

Total 1682 subdivisions.

May find and replace/mark up some popular terms.

Find some popular terms:

Subdivisions in column A,

In B1, add =COUNTIF(\$A:\$A,A1)

Copy down

Then sort by column B to find the times a word appears.

Top occurred subdivisions:

United States 200

Soviet Union 90

Socialism 69

Communism 52

History 35

Politics and government 34

Communist Party of the United States of America 25

Disarmament 20

Congresses 19

Marx, Karl 1818-1883 18

Peace 17

1917-1945 15

MODS Editing: Identify Pattern and Mark Up Data (Subject, cont.)

○ Examples

Replace:

```
<subject>  
  <topic>Labor -- United States</topic>  
</subject>
```

With:

```
<subject><topic>Labor</topic><geographic>United States</geographic></subject>
```

Replace:

```
<subject>  
  <topic>Labor unions -- Great Britain -- History</topic>  
</subject>
```

With:

```
<subject><topic>Labor unions</topic><geographic>Great  
Britain</geographic><topic>History</topic></subject>
```

Replace:

```
<subject>  
  <topic>Soviet Union -- Economic conditions -- 1917-1945</topic>  
</subject>
```

With:

```
<subject><geographic>Soviet Union</geographic><topic>Economic conditions</topic><temporal>1917-  
1945</temporal></subject>
```

Some batch replacement can be done to top occurred subdivisions, but comprehensive review of all subdivisions and manual editing will be needed!!

MODS Editing: Related Item, Collection Title and Physical Description

○ Add related item, e.g.

```
<relatedItem type="original">
  <location>
    <url note="(University Libraries Online
Catalog)">http://ucf.catalog.fcla.edu/permalink.jsp?29CF000388014</url>
  </location>
  <titleInfo>
<title>title.</title>
  </titleInfo>
</relatedItem>
```

(note: need to edit the related item title information)

○ Add collection title, e.g.

```
<relatedItem type="series">
  <titleInfo type="uniform">
    <title>PRISM: Political & Rights Issues & Social Movements collection</title>
  </titleInfo>
  <location>
    <url displayLabel="(Link to Collection)">https://ucf.digital.flvc.org/islandora/object/ucf%3Aasmc</url>
  </location>
</relatedItem>
```

○ Physical description in note field:

- Change <note>47 p. ; 22 cm.</note> to:

```
<physicalDescription>
  <extent>47 p. ; 22 cm.</extent>
</physicalDescription>
```

MODS Editing: File Example

```
<?xml version="1.0" encoding="UTF-8"?>
- <mods version="3.4" xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-4.xsd"
  xmlns:dcterms="http://purl.org/dc/terms/" xmlns:flvc="info:flvc/manifest/v1" xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.loc.gov/mods/v3">
  - <titleInfo>
    <nonSort>The </nonSort>
    <title>meaning of the Soviet trials</title>
  </titleInfo>
  <identifier type="digitool"> 671335 </identifier>
  - <name type="personal">
    <namePart>Yaroslavskii, Emel'ian</namePart>
    <namePart type="date">1878-1943</namePart>
    - <role>
      <roleTerm type="text" authority="marcrelator">creator</roleTerm>
    </role>
  </name>
  - <name type="personal">
    <namePart>Bukharin, Nikolai Ivanovich</namePart>
    <namePart type="date">1888-1938</namePart>
    <role>
      <roleTerm type="text" authority="marcrelator">defendant</roleTerm>
    </role>
    - <role>
      <roleTerm type="text" authority="marcrelator">contributor</roleTerm>
    </role>
  </name>
  . . .
  - <originInfo>
    <publisher>Workers Library Publishers</publisher>
    <dateIssued keyDate="yes" encoding="w3cdtf">1938</dateIssued>
    - <place>
      <placeTerm type="text">New York, N.Y.</placeTerm>
    </place>
  </originInfo>
  - <language>
    <languageTerm type="text">English</languageTerm>
  </language>
  - <physicalDescription>
    <extent>46, [1] p. ; 19 cm.</extent>
  </physicalDescription>
  - <subject>
    - <name type="personal">
      <namePart>Bukharin, Nikolai Ivanovich</namePart>
      <namePart type="date">1888-1938</namePart>
    </name>
    <topic>Trials, litigation, etc</topic>
  </subject>
  - <subject>
    <topic>Trials (Political crimes and offenses)</topic>
    <geographic>Russia (Federation)</geographic>
    <geographic>Moscow</geographic>
  </subject>
```

MODS Editing: File Example (Cont.)

```
- <relatedItem type="original">
  - <location>
    <url note="(Link to Original Online Catalog Record)">http://ucf.catalog.fcla.edu/permalink.jsp?29CF000388014</url>
  </location>
  - <titleInfo>
    <nonSort>The </nonSort>
    <title>meaning of the Soviet trials</title>
  </titleInfo>
</relatedItem>
- <relatedItem type="series">
  - <titleInfo type="uniform">
    <title>PRISM: Political & Rights Issues & Social Movements collection</title>
  </titleInfo>
  - <location>
    <url displayLabel="(Link to PRISM Collection)">https://ucf.digital.flvc.org/islandora/object/ucf%3Aasmc</url>
  </location>
</relatedItem>
- <location>
  <physicalLocation>UCF Libraries Special Collections - 5th Floor -- HX15.V35 no.829</physicalLocation>
</location>
- <location displayLabel="purl">
  <url>http://purl.flvc.org/FCLA/DT/671335</url>
</location>
<accessCondition>All rights to images are held by the respective holding institution. This image is posted publicly for non-profit educational uses, excluding printed publication. For permission to reproduce images and/or for copyright information contact Special Collections and University Archives, University of Central Florida Libraries, (407) 823-2576. http://library.ucf.edu/SpecialCollections/</accessCondition>
<typeOfResource>text</typeOfResource>
</mods>
```

A Resulted MARC Display (only to logged in users)



SEARCH IN

PRISM Collection

for

Go

Home

Advanced Search

View All Items

My account

Embargoed Objects

Log out

View

IP Embargo

Manage

MARCXML

[UCF Digital Collections](#) » [PRISM Collection](#)

[Download MARCXML](#)

```
Leader  ^.....nam^^22^^.....uu^4500
008    ^.....s1938^^.....||||||||||||||||eng||
035    ## |a (digitool) 671335
035    ## |a (IID)CFDT671335
035    ## |a (fedora)ucf:5550
100    1# |a Yaroslavskii, Emel'ian, |d 1878-1943
245    14 |a The meaning of the Soviet trials.
260    ## |a New York, N.Y. : |b Workers Library Publishers, |c 1938
300    ## |a 46, [1] p. ; 19 cm.
380    #7 |a book
534    ## |t meaning of the Soviet trials || http://ucf.catalog.fcla.edu/permalink.jsp?29CF000388014
540    ## |a All rights to images are held by the respective holding institution. This image is posted publicly for
non-profit educational uses, excluding printed publication. For permission to reproduce images and/or for
copyright information contact Special Collections and University Archives, University of Central Florida Libraries,
(407) 823-2576. http://library.ucf.edu/SpecialCollections/
600    14 |a Bukharin, Nikolai Ivanovich |d 1888-1938
650    14 |a Trials (Political crimes and offenses) |z Russia (Federation) |z Moscow
700    1# |a Bukharin, Nikolai Ivanovich, |d 1888-1938, |e defendant
700    1# |a Foster, William Z., |d 1881-1961, |e contributor
700    1# |a Iãgoda, Genrikh Grigorévich, |d 1891-1938, |e defendant
700    1# |a Rykov, Aleksei Ivanovich, |d 1881-1938, |e defendant
830    ## |a PRISM: Political & Rights Issues & Social Movements collection.
852    ## |a UCF Libraries Special Collections - 5th Floor -- HX15.V35 no.829
856    40 |u http://purl.flvc.org/FCLA/DT/671335
887    ## |a owningInstitution="UCF", submittingInstitution="UCF", source="digitool", admin_unit="FCL01",
ingest_id="ing5109", creator="creator:CBILODEAU", creation_date="2010-03-04 09:48:01",
modified_by="creator:CBILODEAU", modification_date="2010-03-04 09:48:18"
998    77 |b text
```

Record Public Display

The meaning of the Soviet trials | ucf.digital.flvc.org - Mozilla Firefox

ucf.digital.flvc.org/islandora/object/ucf%3A5550

University of Central Florida Libraries

SEARCH IN PRISM Collection for [] Go

Home Advanced Search View All Items

UCF Digital Collections » PRISM Collection

The meaning of the Soviet trials

Download pdf | Full Screen View

Summary Full Description

Title: The meaning of the Soviet trials.

Name(s): [Yaroslavskii, Emel'ian, 1878-1943](#), creator
[Bukharin, Nikolai Ivanovich, 1888-1938](#), defendant, contributor
[Foster, William Z., 1881-1961](#), contributor
[Tšogoda, Genrikh Grigorévich, 1891-1938](#), defendant, contributor
[Rykov, Aleksei Ivanovich, 1881-1938](#), defendant, contributor

Type of Resource: text

Genre: book

Date Issued: 1938

Publisher: Workers Library Publishers

Place of Publication: New York, N.Y.

Extent: 46, [1] p. ; 19 cm.

Language: English

Identifier: 671335 (digitool), CFDT671335 (IID), ucf:5550 (fedora)

Subject(s): [Bukharin, Nikolai Ivanovich, 1888-1938](#)
[Trials \(Political crimes and offenses\) -- Russia \(Federation\) -- Moscow](#)

Held by: UCF Libraries Special Collections - 5th Floor -- HX15.V35 no.829

Persistent Link to This Record: <http://purl.flvc.org/FCLA/DT/671335>

Restrictions on Access: All rights to images are held by the respective holding institution. This image is posted publicly for non-profit educational uses, excluding printed publication. For permission to reproduce images and/or for copyright information contact Special Collections and University Archives, University of Central Florida Libraries, (407) 823-2576. <http://library.ucf.edu/SpecialCollections/>

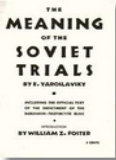
Owner Institution: UCF

Original Title: The meaning of the Soviet trials.
<http://ucf.catalog.fcla.edu/permalink.jsp?29CF000388014> (Link to Original Online Catalog Record)

Is Part of Series: PRISM: Political & Rights Issues & Social Movements collection.
([Link to PRISM Collection](#))

In Collections

- [PRISM Collection](#)



Metadata Migration Issues

- Is there another way to improve the metadata when migrating from a less granular schema to a more granular one besides customizing the XSLT stylesheet and performing post-conversion pattern based records editing? What are the advantages and disadvantages of these methods?
- How to determine the extensibility a stylesheet can be customized prior to post transformation editing in a project?
- Should data mark-up based on patterns and conditions be dealt with in the stylesheet or the post-transformation editing? Which factors affect the decision?
- Does MARCXML data play an important role in the non-Integrated Library System (ILS) environment? Can auto-generated pseudo MARC records be accepted in a CMS or DAMS? To which degree should the converted MODS and MARC records be edited?
- In mapping local elements to MODS, how to determine whether a MODS extension is considered good practice?
- How much automation will a MODS editing tool allow? How much manual or semi-manual editing work is realistic for librarians? What skills are required to perform the task?

Final Thoughts

- There might not be an easy way to get very good data in migrating from a less granular schema to a more granular one. To which degree and in which method should or can the data be improved may depend on many factors.
- The decisions on metadata migration will be different depending on whether the project is on national, consortia or library level.
- Extensive stylesheet customization seems to be more attainable on the collection level when comparable data is available.
- Balance needs to be sought in pre-transformation stylesheet modification and post-transformation records editing.
- It still seems relevant to share collections in a non-traditional DAMS with a traditional ILS in the current library environment.
- Medium or large scale text markup and encoding presents new challenges and requirements for librarians especially cataloging and metadata librarians.

Project Team and Acknowledgments

- Sai Deng, Metadata Librarian, University of Central Florida, sai.deng@ucf.edu
- Lee Dotson, Digital Initiatives Librarian, University of Central Florida, lee.dotson@ucf.edu
- Mary Page, Associate Director, Collections & Technical Services

- Special thanks go to FLVC folks. Inquires can be sent to FLVC Help Desk, help@flvc.org

Thank you!