

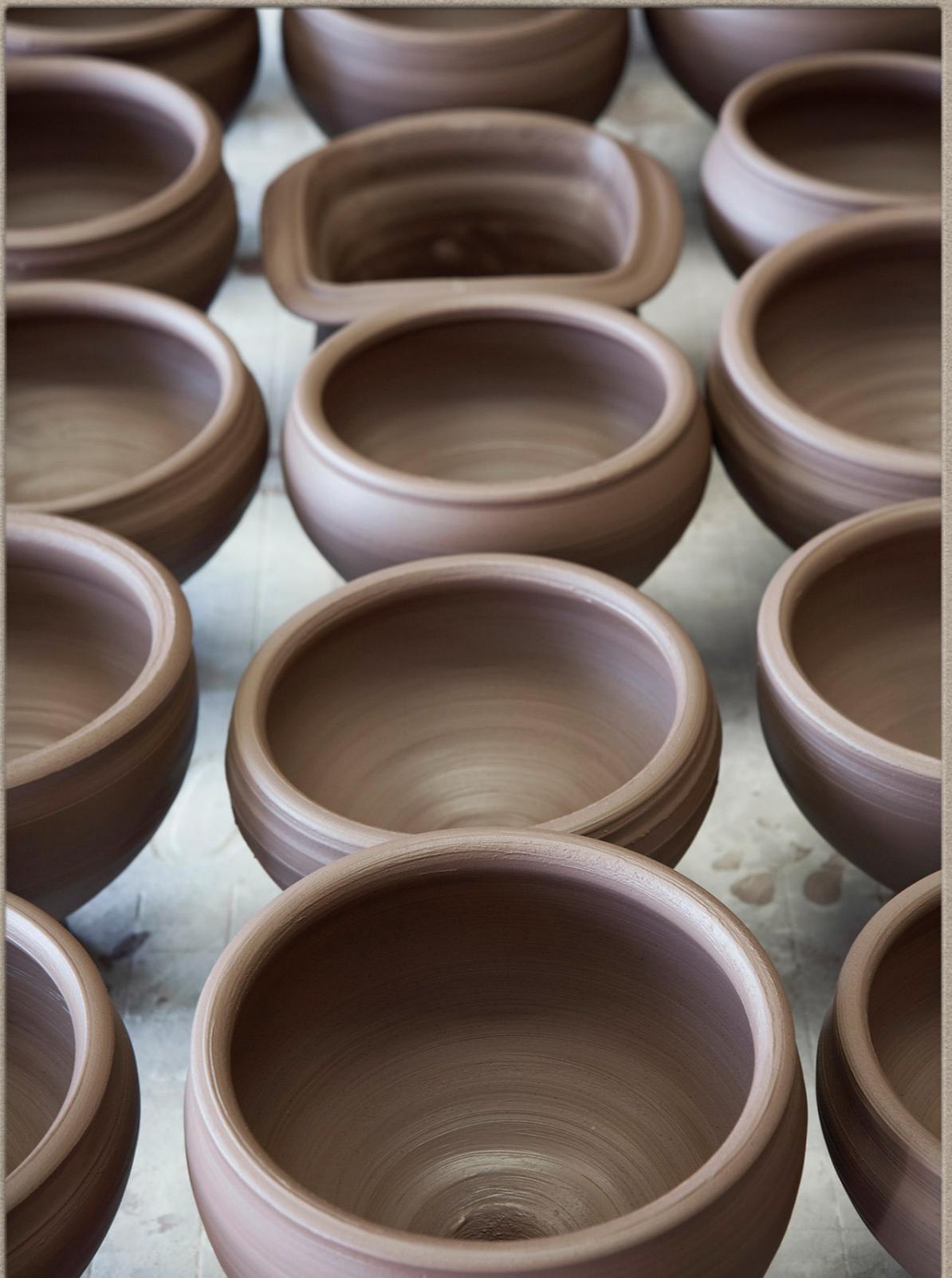
REMEDIATION OF NEAR-MATCH DATA PROCESSING BIBLIOGRAPHIC RECORDS FOR MIGRATION TO A NEW ILS

github.com/aglerum/premigration_remediation

PREAMBLE

WHO WE ARE

WHAT WE ARE
DOING





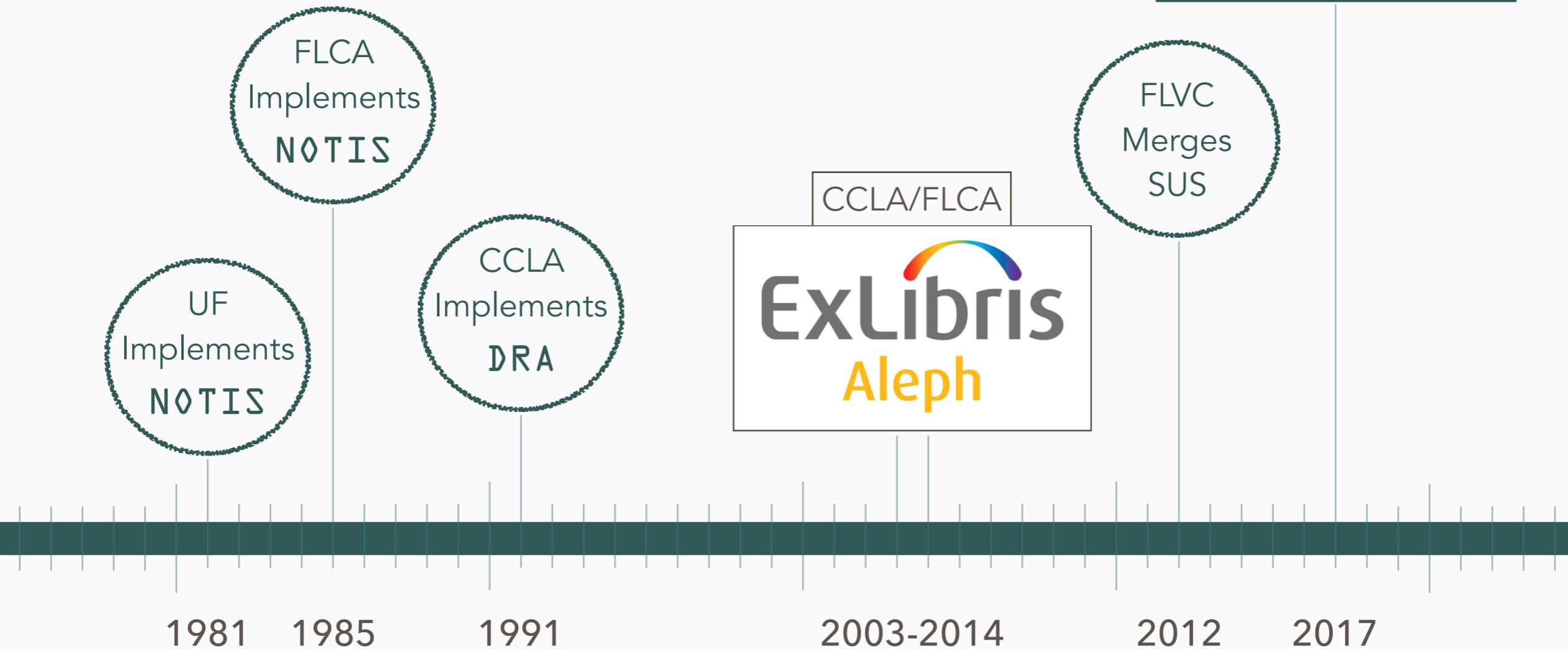
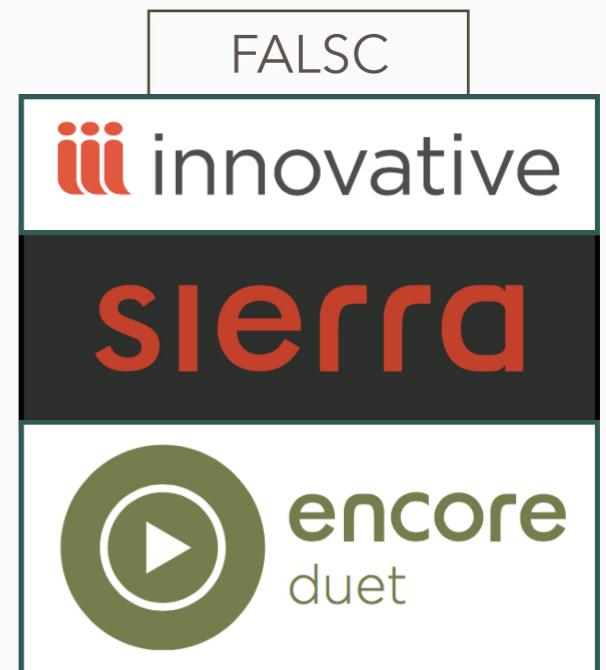
Florida Academic Library Services Cooperative

Services and Information for Florida's College and University Libraries

<https://libraries.flvc.org/about-falsc>

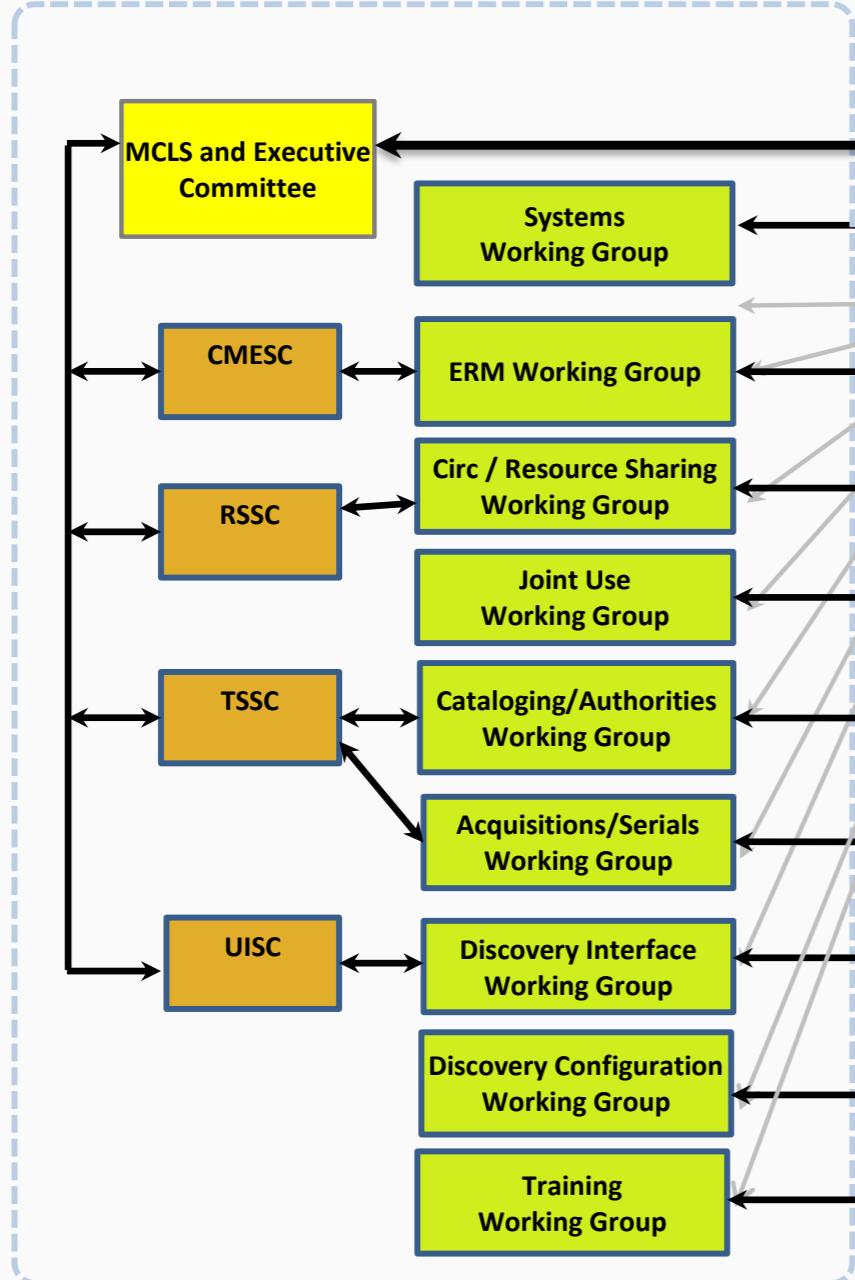


Florida University and Colleges Integrated Library Systems



Organizational Framework for ILS Implementation - Graphical Version

Advisory Groups

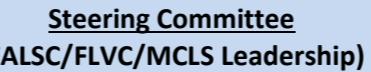


FALSC / FLVC

Steering Committee
(FALSC/FLVC/MCLS Leadership)

IMPLEMENTATION TEAM
Project Managers (2)
FALSC Exe Dir (1)
FLVC staff (9)
Chairs of WG (9)
FCS and SUS rep (2)

TECHNICAL TEAM
All relevant FLVC staff



Institutions

Institutional Leadership
(CIO, Provost, VP,
Deans/Directors)

40 Local ILS
Coordinators

Local
Implementation
Teams

Systems
Discussion Group

ERM Discussion Group

Circ / Resource Sharing
Discussion Group

Joint Use
Discussion Group

Cataloging/Authorities
Discussion Group

Acquisitions/Serials
Discussion Group

Discovery Interface
Discussion Group

Discovery Configuration
Discussion Group

Training Discussion
Group

ISSUES

WHAT THEY ARE

WHY THEY ARE



KEY: Fields that merged

4901 \$a S. hrg. ; \$v 110-432

4901 \$a S. hrg. ; \$v 110-432 \$5 FBoU

**500 \$a Distributed to some depository
libraries in microfiche.**

**500 \$a Distributed to some depository
libraries in microfiche. \$5 FBoU**

500 \$a Shipping list no.: 2008-0423-P.

**500 \$a Shipping list no.: 2008-0423-P. \$5
FBoU**

Fields to Protect During SUS Merge

TAG	NAME
260:2#:	Publication, distribution, etc.
260:3#:	Publication, distribution, etc.
321	Former publication frequency
351	Organization and arrangement of materials
362	Dates of publication and/or sequential designation
400	Series
410	Series
411	Series
440	Series
490	Series
500	General note
501	With note
505 00	Formatted contents note--Enhanced
505 0_	Formatted contents note--Basic
506	Restrictions on access note
515	Numbering peculiarities note

TAG	NAME
516	Type of computer file or data note
518	Date/time and place of an event note
530	Additional physical form available
533	Reproduction note
534	Original version note
538	Systems details note
540	Terms governing use and reproduction note
541	Immediate source of acquisition note
542	Information relating to copyright status
545	Biographical or historical data
552	Entity and attribute information note
561	Ownership and custodial history
562	Copy and version identification note
563	Binding information
583	Action note
584	Accumulation and frequency of use note

TAG	NAME
590	Local note
690	Local topical term
691	Local geographic name
699	Local subject added entry--Uniform title
780	Preceding entry
785	Succeeding entry
790	Local added entry-Personal name
791	Local added entry-Corporate name
796	Local added entry-Personal name
797	Local added entry-Corporate name
845	Terms governing use and reproduction
856	Electronic location and access
856 #2	Electronic location and access
909	Local data elements
951	Local data elements
970	Local data elements

Fields Currently Protected

TAG	NAME
260:2#:	Publication, distribution, etc.
260:3#:	Publication, distribution, etc.
321	Former publication frequency
351	Organization and arrangement of materials
362	Dates of publication and/or sequential designation
400	Series
410	Series
411	Series
440	Series
490	Series
500	General note
501	With note
505 00	Formatted contents note--Enhanced
505 0_	Formatted contents note--Basic
506	Restrictions on access note
515	Numbering peculiarities note

TAG	NAME
516	Type of computer file or data note
518	Date/time and place of an event note
530	Additional physical form available
533	Reproduction note
534	Original version note
538	Systems details note
540	Terms governing use and reproduction note
541	Immediate source of acquisition note
542	Information relating to copyright status
545	Biographical or historical data
552	Entity and attribute information note
561	Ownership and custodial history
562	Copy and version identification note
563	Binding information
583	Action note
584	Accumulation and frequency of use note

TAG	NAME
590	Local note
690	Local topical term
691	Local geographic name
699	Local subject added entry--Uniform title
780	Preceding entry
785	Succeeding entry
790	Local added entry-Personal name
791	Local added entry-Corporate name
796	Local added entry-Personal name
797	Local added entry-Corporate name
845	Terms governing use and reproduction
856	Electronic location and access
856 #2	Electronic location and access
909	Local data elements
951	Local data elements
970	Local data elements

Remove \$5 Subfield for Deduplication

TAG	NAME
260:2#:	Publication, distribution, etc.
260:3#:	Publication, distribution, etc.
321	Former publication frequency
351	Organization and arrangement of materials
362	Dates of publication and/or sequential designation
400	Series
410	Series
411	Series
440	Series
490	Series
500	General note??
501	With note
505 00	Formatted contents note-- Enhanced
505 0_	Formatted contents note--Basic
506	Restrictions on access note
515	Numbering peculiarities note

TAG	NAME
516	Type of computer file or data note
518	Date/time and place of an event note
530	Additional physical form available
533	Reproduction note
534	Original version note
538	Systems details note
540	Terms governing use and reproduction note
541	Immediate source of acquisition note
542	Information relating to copyright status
545	Biographical or historical data
552	Entity and attribute information note
561	Ownership and custodial history
562	Copy and version identification note
563	Binding information
583	Action note
584	Accumulation and frequency of use note

TAG	NAME
590	Local note
690	Local topical term
691	Local geographic name
699	Local subject added entry--Uniform title
780	Preceding entry
785	Succeeding entry
790	Local added entry-Personal name
791	Local added entry-Corporate name
796	Local added entry-Personal name
797	Local added entry-Corporate name
845	Terms governing use and reproduction
856	Electronic location and access
856 #2	Electronic location and access
909	Local data elements
951	Local data elements
970	Local data elements

SOLUTIONS

LOCAL NOTES

NEAR-MATCH DATA



OVERALL WORKFLOW

- (1) Create an element set from MARC data.
- (2) Review and remediate in OpenRefine:
 - Cluster and merge near-match data.
 - Identify local notes in 500 fields.
- (3) Export OpenRefine data to create MARC.

CREATE ELEMENTS SET

Using oXygen

1. Run *create_elements.xsl*.

The result creates brief XML records for:

- Bib Record System Number
- MARC 500 or 590 Field Tag
- MARC Field \$a Text
- MARC Field's \$5 Subfield Code

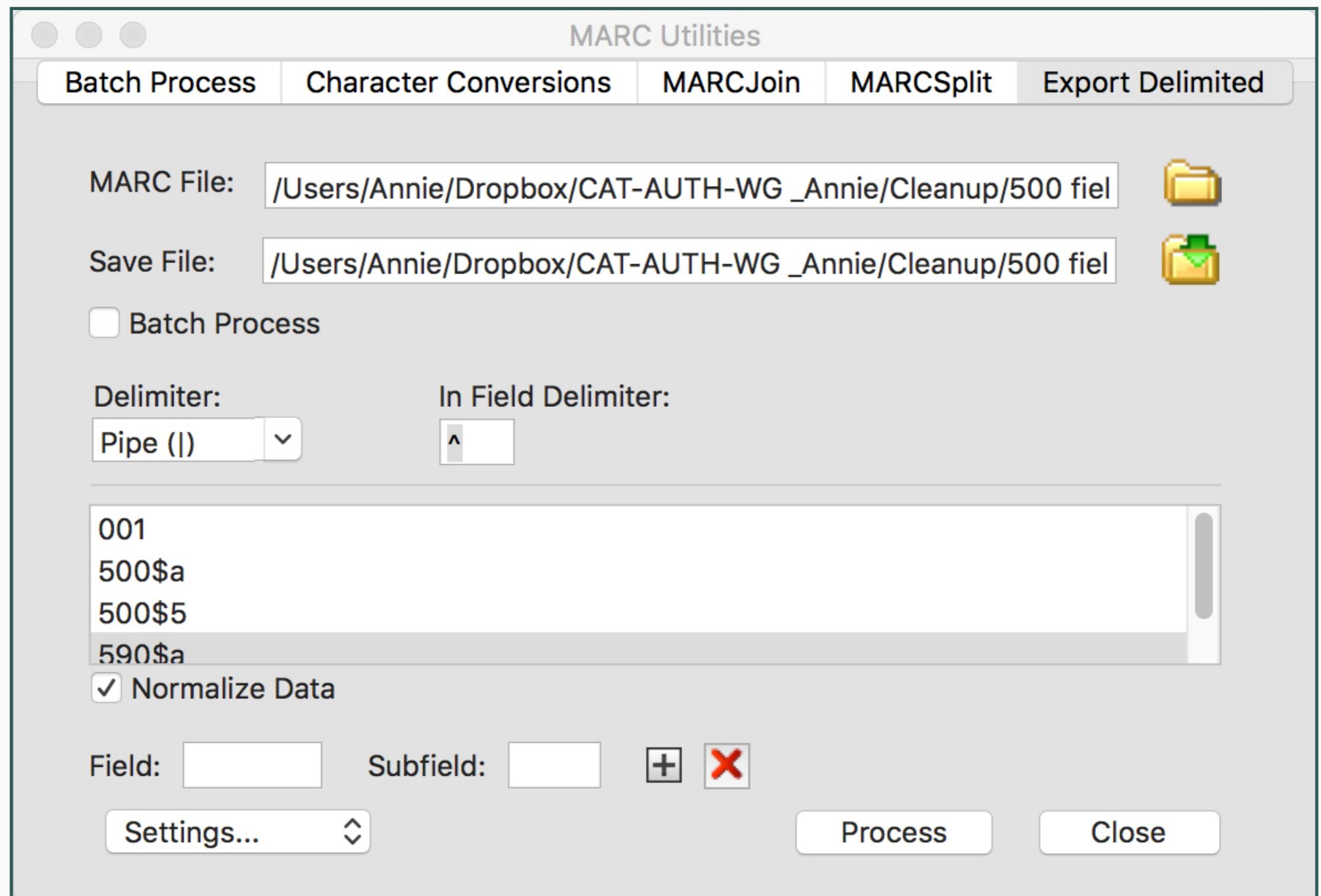
ELEMENTS SET

```
<record>
  <bib>032086172</bib>
  <field>
    <tag>500</tag>
    <text>Title from title screen (viewed Mar. 4,
2009).</text>
    <codes>
      <FMFIU>FIU Library</FMFIU>
    </codes>
  </field>
</record>
```

CREATE ELEMENTS SET

Why not use MarcEdit?

1. Go to MARC Utilities\Export Delimited.
2. Navigate to the MARC file.
3. Navigate to the Save .txt file.
4. Choose the Field and In Field delimiters.
5. Select 001, 500 \$a, 500 \$5, 590 \$a, 590 \$5.



Pipe Delimited Result

001|500\$a|500\$5|590\$a|590\$5

"020000503"|" Title from cover. Running
title: SPLC report."|" FTaSU FTaSU"||

"020000598"|" ""94-801 EPW."" Cover title.
""January 3, 1996."""|" FTaSU^FU FTaSU^FU
FTaSU^FU"||

"020000626"|" ""95-128 L."" Cover title.
""Updated April 12, 1996."""|" FTaSU^FU
FTaSU^FU FTaSU^FU"||

OPEN IN OPENREFINE – TXT

In OpenRefine

1. Choose Create Project tab.
2. Get data from: This Computer/Choose File.
3. Click Next.
4. Accept the defaults.
5. Click Create Project.

The screenshot shows the OpenRefine web application running at 127.0.0.1. The interface includes a top navigation bar with back, forward, and search icons, and a sidebar on the left with project management options like 'Create Project', 'Open Project', 'Import Project', and 'Language Settings'. A large blue diamond icon is present in the sidebar. The main content area displays a welcome message: 'A power tool for working with messy data.' It highlights the ability to 'Create a project by importing data. What kinds of data files can I import?' and lists supported formats: TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents. It also mentions support for other formats via OpenRefine extensions. Below this, a section titled 'Get data from' offers options: 'This Computer' (selected), 'Web Addresses (URLs)', 'Clipboard', and 'Google Data'. A file selection dialog shows 'fslaw1222_corrected.txt' selected, and a 'Next »' button is visible.

127.0.0.1

Refine OPEN

A power tool for working with messy data.

Create Project

Open Project

Import Project

Language Settings

Version 2.6-rc.2 [TRUNK]

Help

About

Get data from

This Computer

Web Addresses (URLs)

Clipboard

Google Data

Choose Files

Next »

The screenshot shows the Refine web application interface. At the top, there is a header bar with navigation icons (back, forward, search, etc.), a URL field containing "127.0.0.1", and several action buttons (refresh, download, upload, etc.). Below the header is the Refine logo with the word "OPEN" in blue above it, followed by the text "A power tool for working with messy data."

On the left side, there is a sidebar with the following options:

- Create Project
- Open Project
- Import Project
- Language Settings

Below the sidebar is a blue diamond icon and the text "Version 2.6-rc.2 [TRUNK]".

On the right side, the main content area has a title "Parse data as" and a sub-section "CSV / TSV / separator-based files". There is a list of file types:

- Line-based text files
- Fixed-width field text files
- PC-Axis text files
- JSON files
- MARC files

Below this, there are settings for "Character encoding" (with a dropdown menu) and "Columns are separated by" (with radio buttons for "commas (CSV)", "tabs (TSV)", and "custom |" which is selected). There is also a note about escaping special characters with "\".

173208 records

Show as: [rows](#) [records](#)

Show: [5](#) [10](#) [25](#) [50](#) records

[« first](#) [< previous](#) [1 - 10](#)

<input type="checkbox"/> All	<input type="checkbox"/> 001	<input type="checkbox"/> 500\$a	<input type="checkbox"/> 500\$5
		1. 020000081 "Substantially the fall 1966 issue of the Wisconsin law review with substitutions and additions including a new introduction by A. Arthur Schiller." "Substantially the fall 1966 issue of the Wisconsin law review with substitutions and additions including a new introduction by A. Arthur Schiller."	FSsNC^FMFIU FJUNF^FBou FTS^FTaSU FU
		2. 020000394	
		3. 020000503 Title from cover. Running title: SPLC report.	FTaSU FTaSU
		4. 020000598 "94-801 EPW." Cover title. "January 3, 1996."	FTaSU^FU FTaSU^FU FTaSU^FU
		5. 020000626 "95-128 L." Cover title. "Updated April 12, 1996."	FTaSU^FU FTaSU^FU FTaSU^FU
		6. 020000656 "95-354 E." Cover title. "Updated November 17, 1995."	FTaSU^FU FTaSU^FU FTaSU^FU
		7. 020000685 "95-373 E." Cover title. "March 3, 1995."	FTaSU^FU FTaSU^FU FTaSU^FU
		8. 020000717 "95-443 E." Cover title. "Updated December 19, 1995."	FTaSU^FU FTaSU^FU FTaSU^FU
		9. 020000748 "95-452 E." Cover title. "March 29, 1995."	FTaSU^FU FTaSU^FU FTaSU^FU
		10. 020000773 Accompanied by: Manual for teachers (385 p. ; 26 cm.) ; c2008.	

Horizontal data

The \$5 data is not associated with a particular field.

Addendum to presentation: OpenRefine can split multiple value cells then fill down to create vertical data. XSLT is preferred because it creates columns for each \$5 code then swaps codes for branch names.

OPEN IN OPENREFINE – XML

In OpenRefine

1. Choose Create Project tab.
2. Get data from: This Computer/Choose File.
3. Click Next.
4. Select the <record> element.
5. Click Create Project.

« Start Over Configure Parsing Options Project name **fsmed1222_elements xml** **Create Project »**

Click on the first XML element corresponding to the first record to load.

```
<records>
  <record>
    <bib>034366121</bib>
    <field>
      <tag>500</tag>
      <text>Revised edition of: Abdominal imaging / [edited by] Dushyant V. Sahani, Anthony E. Samir. 1st ed. ©2011</text>
      <codes />
    </field>
  </record>
  <record>
    <bib>034719688</bib>
    <field>
```

Parse data as

XML files

[Open Document Format spreadsheets \(.ods\)](#)

[RDF/XML files](#)

[JSON files](#)

[Line-based text files](#)

Load at most 0 row(s) of data

Preserve empty strings

Trim leading & trailing whitespace from strings

Parse cell text into numbers, dates, ...

Store file source
(file names, URLs)
in each row

Pick Record Elements Update Preview

582738 records								Extensions	
Show as: rows records			Show: 5 10 25 50 records			« first < previous 1 - 10 next > last »			
<input type="checkbox"/> All	<input type="checkbox"/> record	<input type="checkbox"/> record - field	<input type="checkbox"/> record - field - c	<input type="checkbox"/> record - field - c	<input type="checkbox"/> record - field - c	<input type="checkbox"/> record - field - c	<input type="checkbox"/> record - field - c	<input type="checkbox"/> record - field - c	
		1.							
		2.							
		3.							
		4.							
				FSU Libraries	UF Library	FIU Library			

XML import has empty columns and rows.

Format the Open Refine data

Note: Formatting is not required, but it easier to identify columns with shorter names. Blank rows should be removed.

1. Open in the "elements" file in OpenRefine (e.g. `fsmed_elements.xml`).
2. Remove columns for the parent elements: `record` , `record - tags` , and `record - tags - codes` .
3. In the **All** column pull down menu, go to **Edit columns>>Re-order / Remove columns....**
4. Move columns to this order: **bib | tag | text | [codes]**.
5. Rename columns to remove all preceding sibling names. e.g. `record - tag - codes - FTaSU` becomes `FTaSU` .
6. In the **bib** column pull down menu, go to **Edit cells>>Fill down**.
7. In the **tag** column pull down menu, go to **Filter>>Text Filter**.
8. In the **Facet/Filter** panel, select **blank**.
9. In the **All** column pull down menu, go to **Edit rows>>Remove all matching rows**.
10. Close the **Facet/Filter** panel.

https://github.com/aglerum/premigration_remediation/wiki/Remediate-Local-Data-II:-Review-and-Remediate-in-OpenRefine

292275 rows

Extensions:

Show as: **rows** records

Show: **5 10 25 50** rows

« first < previous **1 - 10** next > last »

All	bib	tag	text	FBoU	FFmFGC	FIMiFIUC	FJUNF	FMFIU	FOFT	FPe
1.	020000081	500	"Substantially the fall 1966 issue of the Wisconsin law review with substitutions and additions including a new introduction by A. Arthur Schiller."							
2.	020000081	500	"Substantially the fall 1966 issue of the Wisconsin law review with substitutions and additions including a new introduction by A. Arthur Schiller."	FAU Library			UNF Library	FIU Library		
3.	020000503	500	Title from cover.							
4.	020000503	500	Running title: SPLC report.							
5.	020000598	500	"94-801 EPW."							

Vertical data

Each field and its associated \$5 subfield in its own row.

REVIEW & REMEDIATE

Examine 500 fields for possible local data

1. In the **tag** column pull down menu, go to **Filter>>Text Filter**.
2. In the **Facet/Filter** panel, select **500**.
3. In the **tag** column pull down menu, go to **Edit rows>>Cluster and edit**.
4. Choose the most common version of the notes or preferred note.
 - Select any quoted notes with caution. Sometimes only difference is incorrect punctuation or spacing (e.g., note missing period at end or an extra or missing space). For RDA, capitalization does matter, however, okay to make local or consortial decision on standardizing capitalization in notes. For example, there may be instances of **500 __ "July 12, 2003."** and **500 __ July 12, 2016.** These could be standardized as a recorded note (**500 __ July 12, 2016.**). However, in the case of a series-like phrase, the note should be transcribed and quoted. To determine if these similar notes are found in the same record, click **Browse this Cluster** at the bottom of that text group. This option only appears if the selection box is unchecked.
 - Occasionally, there will situations where the same data is present in multiple 500 fields and one has a \$5 and the other one doesn't. Search OCLC to confirm if this is truly local data. For example, "Author's autograph presentation copy. Signed." looks like a local note, but the record is for an online version, and since this is true for the digitized copy, it is not a local note.

REVIEW & REMEDIATE

Examine 500 fields for possible local data

5. Correct any errors, such as **Title from BioMed Central archive volume screen (viewed Mar. 9, 2005)** to **...(viewed Mar. 9, 2005)**.
6. Click **Merge selected and Close**.
7. In the **tag** column pull down menu, go to **Filter>>Text Filter**.
8. In the **Facet/Filter** panel, select **500** then **Sort by count**.
9. Scan the list to identify possible local notes. e.g. **Donated by Dr. Charlotte Maguire.** with only one code -- **FtasU**.
10. If the note is determined to be local:
 - i. In the **Facet/Filter** panel, select the local note.
 - ii. In the **tag** column, select the cell of the first row, click **Edit**, enter **590**, then **Apply to All Identical Cells**.
11. If the note doesn't have an institutional prefix, select the cell of the first row, click **Edit**, enter the prefix, then **Apply to All Identical Cells**.
12. If the note indicates an institution, but there isn't a \$5 code, add the Institution Branch (i.e. **FSU Libraries**) in the column for the appropriate MARC Organization Code.

Cluster & Edit column "marc500"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method key collision ▾

Keying Function fingerprint ▾

345 clusters found

Cluster Size	Row Count	Values in Cluster	Merge? <input type="checkbox"/>	New Cell Value
12	57	<ul style="list-style-type: none"> Date, place of publication and publishers' names from Wing. (18 rows) Place, date of publication, and publisher's names from Wing. (14 rows) Date, place of publication and publisher's names from Wing. (12 rows) Place and date of publication and publisher's names from Wing. (3 rows) Publishers' names, date and place of publication from Wing. (3 rows) Date and place of publication, publishers' names from Wing. (1 rows) Date, place of publication and publishers' names from Wing.. (1 rows) Date, place of publication, and publishers' names from Wing. (1 rows) Place, date of publication and publisher's names from Wing. (1 rows) Place, date of publication and publishers' names from Wing. (1 rows) Publisher's names, place and date of publication from Wing. (1 rows) Publishers' names, place and date of publication from Wing. (1 rows) 	<input type="checkbox"/>	Date, place of publication and p
9	102	<ul style="list-style-type: none"> Date, place of publication and publisher's name from 	<input type="checkbox"/>	Date, place of publication and p

Select All
Unselect All
Merge Selected & Re-Cluster
Merge Selected & Close
Close

Choices in Cluster

2 — 12

Rows in Cluster

0 — 2100

Average Length of Choices

0 — 1000

Length Variance of Choices

0 — 15

Cluster Size	Row Count	Values in Cluster	Merge?
12	57	<ul style="list-style-type: none"> • Date, place of publication and publishers' names from Wing. (18 rows) • Place, date of publication, and publisher's names from Wing. (14 rows) • Date, place of publication and publisher's names from Wing. (12 rows) • Place and date of publication and publisher's names from Wing. (3 rows) • Publishers' names, date and place of publication from Wing. (3 rows) • Date and place of publication, publishers' names from Wing. (1 rows) • Date, place of publication and publishers' names from Wing.. (1 rows) • Date, place of publication, and publishers' names from Wing. (1 rows) • Place, date of publication and publisher's names from Wing. (1 rows) • Place, date of publication and publishers' names from Wing. (1 rows) 	<input type="checkbox"/>

EXPORT PROJECT

In OpenRefine

1. Complete editing the data.
2. Go to the Export dropdown menu.
3. Choose Templating....
4. Replace the default template with your own.
5. Click Export.

OpenRefine Edits Complete

14158 matching rows (292275 total)

Show as: **rows** records

Show: 5 10 25 50 rows

« first < previous 1 - 10

<input type="checkbox"/> All	<input type="checkbox"/> bib	<input type="checkbox"/> tag	<input type="checkbox"/> text	<input type="checkbox"/> FBoU	<input type="checkbox"/> FFmFGC	<input type="checkbox"/> FIMiFIUC	<input type="checkbox"/> FJUNF
		109.	020001649	590	FSU Libraries: Library also has the guide to this microfiche collection, found under: MICRO LAW KF4545 S5 A127 1989.		
		119.	020001690	590	FIU Library: TITLE CANCELLED 11/24/2004		
		165.	020002262	590	UNF Copy : Florida Blue Corporate Library.		
		381.	020005395	590	UNF Library: Reprinted, 1969.		UNF Library
		465.	020006496	590	UWF Library: 75-5309		
		620.	020008180	590	FSU Libraries: Florida State University faculty publication.		
		674.	020008800	590	UNF Library: Second printing, 1985.		UNF Library
		768.	020009779	590	UNF Library: C.2 sixth printing, 1970.		UNF Library
		927.	020012070	590	FAMU Library: SIGNED BY AUTHOR		
		1934.	020027361	590	FSU Libraries: Latest edition only in Reference.		

OpenRefine Default Template

Templating Export

Prefix

```
{  
  "rows" : [
```

Row Template

```
{  
  "bib" : {{jsonize(cells["bib"].value)}},  
  "tag" : {{jsonize(cells["tag"].value)}},  
  "text" : {{jsonize(cells["text"].value)}},  
  "FBoU" : {{jsonize(cells["FBoU"].value)}},  
  "FFmFGC" : {{jsonize(cells["FFmFGC"].value)}},  
  "FlMiFIUC" : {{jsonize(cells["FlMiFIUC"].value)}},  
  "FJUNF" : {{jsonize(cells["FJUNF"].value)}},  
  "FMFIU" : {{jsonize(cells["FMFIU"].value)}},  
  "FOFT" : {{jsonize(cells["FOFT"].value)}},  
  "FPeU" : {{jsonize(cells["FPeU"].value)}},  
  "FSsNC" : {{jsonize(cells["FSsNC"].value)}},  
  "FTaFA" : {{jsonize(cells["FTaFA"].value)}},  
  "FTaSU" : {{jsonize(cells["FTaSU"].value)}},  
  "FTaSU-L" : {{jsonize(cells["FTaSU-L"].value)}},  
  "FTS" : {{jsonize(cells["FTS"].value)}},  
  "FU" : {{jsonize(cells["FU"].value)}}},
```

Row Separator

```
,
```

Suffix

```
]  
}
```

```
{  
  "rows" : [  
    {  
      "bib" : "020000081",  
      "tag" : "500",  
      "text" : "\"Substantially the fall 1966 is  
      "FBoU" : null,  
      "FFmFGC" : null,  
      "FlMiFIUC" : null,  
      "FJUNF" : null,  
      "FMFIU" : null,  
      "FOFT" : null,  
      "FPeU" : null,  
      "FSsNC" : null,  
      "FTaFA" : null,  
      "FTaSU" : null,  
      "FTaSU-L" : null,  
      "FTS" : null,  
      "FU" : null,  
      "FU-L" : null  
    },  
    {  
      "bib" : "020000081",  
      "tag" : "500",  
      "text" : "\"Substantially the fall 1966 is  
      "FBoU" : "FAU Library",  
      "FFmFGC" : null,  
      "FlMiFIUC" : null,  
      "FJUNF" : "UNF Library",  
      "FMFIU" : "FIU Library",  
      "FOFT" : null,  
      "FPeU" : null,  
      "FSsNC" : "NCF Library",  
    }]
```

OpenRefine Default Template

```
{  
  "bib" : {{jsonize(cells["bib"].value)}} ,
```

OpenRefine MARC Template

Templating Export

Prefix

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns:marc="http://www.loc.gov/MARC21/slim"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.loc.gov/MARC21/slim
    http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
```

Row Template

```
<record>
<leader>00000cam 2200000 4500</leader>
<controlfield tag="001">{{cells["bib"].value}}</controlfield>
<datafield tag="245" ind1="0" ind2="0">
  <subfield code="a">TITLE</subfield>
</datafield>
<datafield tag="{{cells["tag"].value}}" ind1=" " ind2=" ">
  <subfield code="a">{{cells["text"].value}}</subfield>
  <subfield code="5">{{cells["FBoU"].value}}</subfield>
  <subfield code="5">{{cells["FFmFGC"].value}}</subfield>
  <subfield code="5">{{cells["FJUNF"].value}}</subfield>
  <subfield code="5">{{cells["FMFIU"].value}}</subfield>
  <subfield code="5">{{cells["FOFT"].value}}</subfield>
  <subfield code="5">{{cells["FPeU"].value}}</subfield>
  <subfield code="5">{{cells["FTaFA"].value}}</subfield>
  <subfield code="5">{{cells["FTaSU"].value}}</subfield>
  <subfield code="5">{{cells["FTS"].value}}</subfield>
  <subfield code="5">{{cells["FU"].value}}</subfield>
</datafield>
</record>
```

Row Separator

```
//
```

Suffix

```
</collection>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns:marc="http://www.loc.gov/MARC21/
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-in
  xsi:schemaLocation="http://www.loc.gov/MARC21/
    http://www.loc.gov/standards/marcxml/schema/MA
<leader>00000cam 2200000 4500</leader>
<controlfield tag="001">033389919</controlfield>
<datafield tag="245" ind1="0" ind2="0">
  <subfield code="a">TITLE</subfield>
</datafield>
<datafield tag="590" ind1=" " ind2=" ">
  <subfield code="a">Elsevier Science</subfield>
  <subfield code="5">null</subfield>
  <subfield code="5">null</subfield>
</datafield>
</record><record>
<leader>00000cam 2200000 4500</leader>
<controlfield tag="001">033405978</controlfield>
<datafield tag="245" ind1="0" ind2="0">
  <subfield code="a">TITLE</subfield>
</datafield>
<datafield tag="590" ind1=" " ind2=" ">
  <subfield code="a">Elsevier</subfield>
  <subfield code="5">null</subfield>
  <subfield code="5">null</subfield>
```

OpenRefine MARC Template

```
<record>
<leader>00000cam 2200000 4500</leader>
<controlfield tag="001">{{cells["bib"].value}}</controlfield>
```

OpenRefine Template Output

```
<record>
  <leader>00000cam 2200000 4500</leader>
  <controlfield tag="001">027392337</controlfield>
  <datafield tag="245" ind1="0" ind2="0">
    <subfield code="a">TITLE</subfield>
  </datafield>
  <datafield tag="590" ind1=" " ind2=" ">
    <subfield code="a">FIU Library: 4th Hubert
      Library copy includes letter.</subfield>
    <subfield code="5">null</subfield>
    <subfield code="5">null</subfield>
    <subfield code="5">FIU Library</subfield>
    <subfield code="5">null</subfield>
  </datafield>
</record>
```

FIX THE TEMPLATE EXPORT

In oXygen

1. Open the template result XML.
2. Run template_fix1.xsl XSLT to remove \$5 subfields with “null”.
3. On that result, run template_fix2.xsl XSLT to replace institution names with institution codes.
4. Save that as the brief record set.

template_fix1.xsl

```
<xsl:stylesheet
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"
    xmlns="http://www.loc.gov/MARC21/slim"
    exclude-result-prefixes="xs xsi"
    version="2.0">
    <xsl:output indent="yes" encoding="UTF-8" method="xml"/></xsl:output>
    <!-- This deletes the space left by deleting the subfields with null. -->
    <xsl:strip-space elements="*"/>
    <xsl:variable name="institutions" select="document('XML/SUS_codes.xml')/institutions/institution"/>
    <!-- Generic identify template -->
    <xsl:template match="@* | node()">
        <xsl:copy>
            <xsl:apply-templates select="@* | node()" />
        </xsl:copy>
    </xsl:template>
    <!-- This deletes the subfields with null. -->
    <xsl:template match="./*[text() = 'null']">
</xsl:stylesheet>
```

Generic Identity Template

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

Delete Any Children with value of ‘null’.

```
<xsl:template match="/*[text() = 'null']"/>
```

```

<xsl:stylesheet
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xmlns:marc="http://www.loc.gov/MARC21/slim"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"
    exclude-result-prefixes="xd xs xsi"
    version="2.0">
    <xsl:output indent="yes" encoding="UTF-8" method="xml"/>
    <xsl:variable name="institutions" select="document('XML/SUS_codes.xml')/institutions/institution"/>
    <!-- Generic identify template -->
    <xsl:template match="@* | node()">
        <xsl:copy>
            <xsl:apply-templates select="@* | node()" />
        </xsl:copy>
    </xsl:template>
    <!-- This swaps the institution name for the institution code in SUS_codes.xml. -->
    <xsl:template match="/*/*/* marc:subfield[@code = '5']">
        <marc:subfield code="5">
            <xsl:value-of select="$institutions[branch = current()]/[(self::code)]"/>
        </marc:subfield>
    </xsl:template>
</xsl:stylesheet>

```

template_fix2.xsl

\$institutions

```
<xsl:variable name="institutions"
select="document('XML/SUS_codes.xml')/institutions/
institution"/>
```

SUS_codes.xml

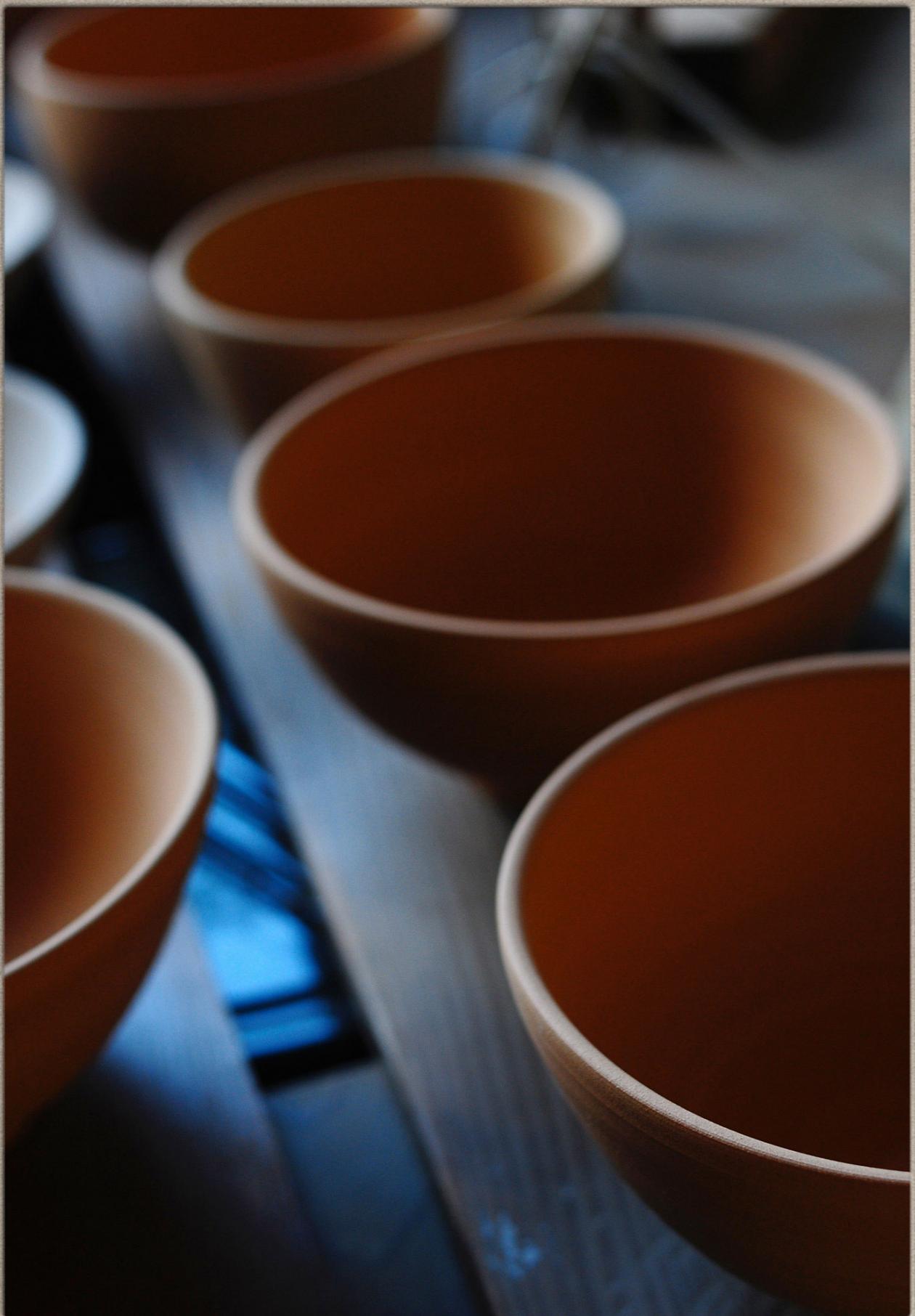
```
<institutions>
  <institution Name = "CRL">
    <branch>Center for Research Libraries (CRL)
    </branch>
    <code>ICRL</code>
  </institution>
[...]
```

Generic Identity Template

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

Swap Institution Name with Code

```
<xsl:template match="/*/*/*/marc:subfield[@code = '5']">
  <marc:subfield code="5">
    <xsl:value-of select=
      $institutions[branch = current()]/*[self::code]" />
  </marc:subfield>
</xsl:template>
```





REMEDIATION OF NEAR-MATCH DATA

Annie Glerum

Head of Complex Cataloging, Florida State University

Email: aglerum@fsu.edu

GitHub: https://github.com/aglerum/premigration_remediation

Website: <http://myweb.fsu.edu/aglerum/index.html>

Twitter: [@annie_glerum](https://twitter.com/annie_glerum)



ACRONYMS

- **CCLA** = Center for College Library Automation
- **CMESC** = Collection Management and E-Resources Standing Committee
- **DRA** = Data Resource Associates
- **FCS** = Florida College System
- **FCLA** = Florida Center for Library Automation
- **FLVC** = Florida Virtual Campus
- **FALSC** = Florida Academic Library Services Cooperative
- **NOTIS** = Northwestern Online Total Integrated System
- **RSSC** = Resource Sharing Standing Committee
- **SUS** = State University System
- **TSSC** = Technical Services Standing Committee
- **UISC** = User Interface Standing Committee

RESOURCES

- **ABOUT SIERRA/ENCORE DUET.** <https://libraries.flvc.org/about-sierra-encore-duet>.
- **Background and Current Status of the Florida CEnter for Library Automation (FCLA) ,College Center for Library Automation (CCLA), SUNLINK and Florida Electronic Library.** <http://dlis.dos.state.fl.us/libraryCouncil/pdfs/Sunlink.pdf>.
- **Directed Technological Change in the Florida Community College System.** <http://www.ala.org/acrl/publications/booksanddigitalresources/booksmonographs/pil/pil49/roark>.
- **FALSC ILS Implementation Organizational Framework.** <https://libraries.flvc.org/documents/181844/388082/Organizational+Framework+for+ILS+Implementation+5-9-16.pdf>.
- **OpenRefine.** <http://openrefine.org>.
- **University of Florida/FCLA and NOTIS.** <http://notis-history.northwestern.edu/Florida.htm>.