

Creating a Discovery Import Profile for Generic XML Source Records in Alma

Presented By: Jason Griffith

eBug Annual Conference 2021

June 14, 2021

Project Purpose

- To display digitized newspapers from the Kentucky Digital Newspaper Project in InfoKat Discovery (Primo).
- Why?
 - Funding
 - Infrastructure
 - Fully use Alma/Primo capabilities:
 - Greater efficiency
 - User needs

KDNP Background

Abstract - Over fifty years ago an historian and a library director traveled the back roads of Kentucky (USA) with a portable microfilm camera, two lights, and a dream of preserving Kentucky's newspapers. From their ambitions arose a successful newspaper preservation program at the University of Kentucky Libraries (UKL). Now in its sixth decade, the program has developed a new way of preserving contemporary born-digital newspapers. This paper explores some of the people and events behind the early success of UKL's program, as well as an in-depth look at the development and functionality of Paper Vault: a largely automated, in-house process delivering and preserving Kentucky's born-digital contemporary newspapers (Terry, Kopana and Weig, Eric, "Preserving Kentucky's Newspapers: Analogue Beginnings to Digital Frontier" (2015). *Library Presentations*. 137. https://uknowledge.uky.edu/libraries_present/137).

KDNP Statistics

- 260+ Kentucky newspaper titles currently available
- Issues dating back to the 18th century
- Thematic representations in collection:
 - African American
 - Emancipation
 - College/University
 - Agriculture
 - Political (Democratic, Independent, Republican, Whig)
 - Civil War Era (Largest assembly in the collection)
- Approx. 87k metadata records

Project Background in Alma

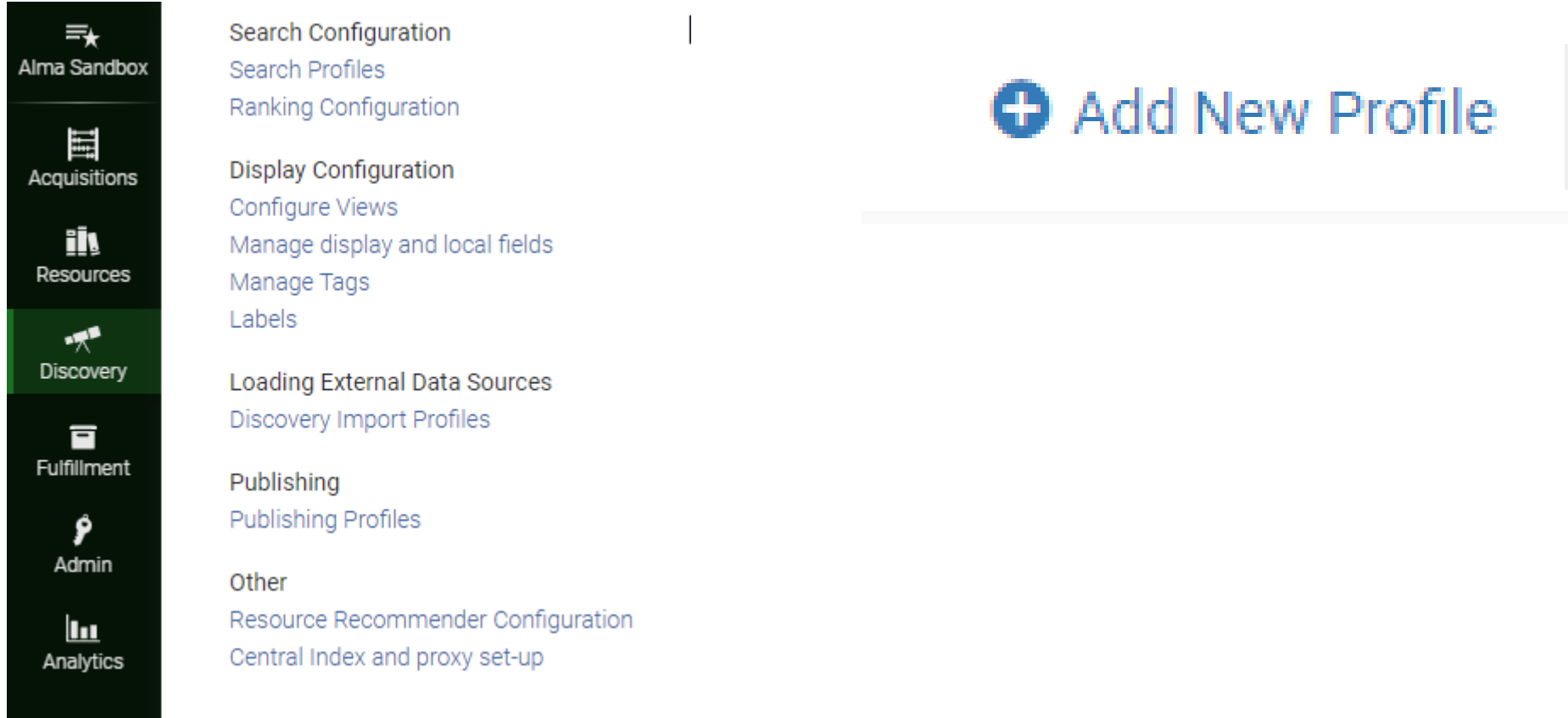
- Questions:
 1. What method of display will we use?
 - Import into Alma and display using Collection Discovery
 - [Use a Discovery Import Profile](#)
 2. Where will the data live?
 - SFTP
 - OAI
 - [Alma](#)
 3. What record format will we use?
 - MARC
 - Dublin Core
 - [Generic XML](#)

Steps Taken for Implementation

1. Create Discovery Import Profile
2. Create Normalization Rules
3. Create Normalization Process
4. Create Local Fields
5. Create Local Resource Type
6. Map Thumbnails for Display

1. Create Discovery Import Profile

- Go to Discovery>Loading External Data Sources>Discovery Import Profiles>Add New Profile



[https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_\(English\)/100Loading_Records_from_External_Sources_into_Primo_VE/Configuring_Import_Profiles_for_Primo_VE](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/100Loading_Records_from_External_Sources_into_Primo_VE/Configuring_Import_Profiles_for_Primo_VE)

Choose Profile Type

- New Order
- Update Inventory
- Repository
- Digital
- Collection
- Discovery

Discovery - Used for loading records from external systems for discovery in Primo VE

Profile Details ⌵

Profile name * **KDNP Newspapers**

Profile description **Import profile for KDNP newspaper issues**

Data Source Code * **KDNP**

Data Source Label * **KDNP** ⓘ

Originating system * **Other** ⌵

File name patterns ⓘ

Import Protocol **Upload File/s** ⓘ

Physical source format **XML** ⓘ

Source format **Generic XML** ⌵

Target format **Dublin Core** ⓘ

Status **Active** ⌵

File splitter parameters
▼

<p>Root element tag * <input style="width: 90%;" type="text"/></p>	<p>Record elements tag * <input style="width: 90%;" type="text"/></p>
<p>XPath to the identifier tag * <input style="width: 90%;" type="text"/></p>	<p>XPath to the location of the deleted status <input style="width: 90%;" type="text"/></p>
<p>Delete record regular expression <input style="width: 90%;" type="text"/></p>	

File Splitter Parameters (Generic XML only)	
Root element tag	The name of the first tag in the file, which should be the name of the collection tag.
Record elements tag	The name of the tag representing the beginning of a record. A file may contain one or more records.
XPath to the identifier tag	The XPath to the identifier tag of the record. This should be the tag that contains the unique identifier for the record.
XPath to the location of the deleted status	The XPath to the location of the deleted status of the record. Default value for a record is: deleted = false .
Delete record regular expression	This regular expression is applied to the value found under the XPath to the location of the deleted status field. If the regular expression matches the value found, the record is marked as deleted.

File Splitter Parameters (Generic XML only)	
Root element tag	The name of the first tag in the file, which should be the name of the collection tag.
Record elements tag	The name of the tag representing the beginning of a record. A file may contain one or more records.
XPath to the identifier tag	The XPath to the identifier tag of the record. This should be the tag that contains the unique identifier for the record.
XPath to the location of the deleted status	The XPath to the location of the deleted status of the record. Default value for a record is: <code>deleted = false</code> .
Delete record regular expression	This regular expression is applied to the value found under the XPath to the location of the deleted status field. If the regular expression matches the value found, the record is marked as deleted.

```

<?xml version="1.0" encoding="UTF-8" standalone="true"?>
<collection>
  <newspaperissue>
    <lccn_link_s>http://chroniclingamerica.loc.gov/lccn/sn86069496/</lccn_link_s>
    <aip_t>xt70rx938d1v</aip_t>
    <id>xt7000000n98</id>
    <lccn_t>sn86069496</lccn_t>
    <printurl_t>https://archive.org/download/xt7000000n98/xt7000000n98.pdf</printurl_t>
    <thumbnailurl_t>https://www.archive.org/download/xt7000000n98/page/n0_w100.jpg</thumbnailurl_t>
    <objecturl_t>https://kentuckynewspapers.org/view.php?id=xt7000000n98</objecturl_t>
    <abbreviation_t>ada</abbreviation_t>
    <control_t>ada1908040801</control_t>
    <edition_s>01</edition_s>
    <collection_s>Adair County news (The)</collection_s>
    <title_t>Adair county news (The)</title_t>
    <region_s>Pennyrile</region_s>
    <doctitle_t>The Adair County news., 1908-04-08</doctitle_t>
    <pubplace_t>Columbia</pubplace_t>
    <publisher_t>Chas. S. Harris</publisher_t>
    <county_s>Adair</county_s>
    <rights_t>http://creativecommons.org/licenses/by/4.0/</rights_t>
    <decade_s>1900</decade_s>
    <year_s>1908</year_s>
    <digital_date_t>2013</digital_date_t>
    <issuedate_t>1908-04-08</issuedate_t>
    <month_s>04</month_s>
    <day_s>08</day_s>
    <pages_s>8</pages_s>
    <funders_t>National Endowment for the Humanities</funders_t>
    <partners_t>Library of Congress</partners_t>
    <format_t>Newspaper issue</format_t>
  </newspaperissue>
</collection>

```

Profile Details

Profile name *	<input type="text" value="KDNP Newspapers"/>
Profile description	<input type="text" value="Import profile for KDNP newspaper issues"/>
Data Source Code *	<input type="text" value="KDNP"/>
Data Source Label *	<input type="text" value="KDNP"/>
Originating system *	<input type="text" value="Other"/>
Import Protocol	<input type="text" value="Upload File/s"/>
Physical source format	<input type="text" value="XML"/>
Source format	<input type="text" value="Generic XML"/>
Target format	<input type="text" value="Dublin Core"/>
Status	<input type="text" value="Active"/>

File splitter parameters

Root element tag *	<input type="text" value="collection"/>
Record elements tag *	<input type="text" value="newspaperissue"/>
XPath to the identifier tag *	<input type="text" value="newspaperissue/id"/>
XPath to the location of the deleted status	<input type="text"/>
Delete record regular expression	<input type="text"/>

2. Create Normalization Rules

- Started off with Generic XML that mimics JSON format used for KDNP records
 - Received file for one newspaper title with 1,170 total issues
- Had to find a method to normalize the Generic XML to DC fields
 - Started with one issue to test

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <collection>
  - <newspaperissue>
    <lccn_link_s>http://chroniclingamerica.loc.gov/lccn/sn86069496/</lccn_link_s>
    <aip_t>xt70rx938d1v</aip_t>
    <id>xt7000000n98</id>
    <lccn_t>sn86069496</lccn_t>
    <printurl_t>https://archive.org/download/xt7000000n98/xt7000000n98.pdf</printurl_t>
    <thumbnailurl_t>https://www.archive.org/download/xt7000000n98/page/n0\_w100.jpg</thumbnailurl_t>
    <objecturl_t>https://kentuckynewspapers.org/view.php?id=xt7000000n98</objecturl_t>
    <abbreviation_t>ada</abbreviation_t>
    <control_t>ada1908040801</control_t>
    <edition_s>01</edition_s>
    <collection_s>Adair County news (The)</collection_s>
    <title_t>Adair county news (The)</title_t>
    <region_s>Pennyrile</region_s>
    <doctitle_t>The Adair County news., 1908-04-08</doctitle_t>
    <pubplace_t>Columbia</pubplace_t>
    <publisher_t>Chas. S. Harris</publisher_t>
    <county_s>Adair</county_s>
    <rights_t>http://creativecommons.org/licenses/by/4.0/</rights_t>
    <decade_s>1900</decade_s>
    <year_s>1908</year_s>
    <digital_date_t>2013</digital_date_t>
    <issuedate_t>1908-04-08</issuedate_t>
    <month_s>04</month_s>
    <day_s>08</day_s>
    <pages_s>8</pages_s>
    <funders_t>National Endowment for the Humanities</funders_t>
    <partners_t>Library of Congress</partners_t>
    <format_t>Newspaper issue</format_t>
  </newspaperissue>
</collection>
```

Create rule in Alma:

Go to Resources>Open Metadata Editor>File>New>Normalization rules (XML)

The image displays two sequential screenshots of the 'Normalization rules (XML) properties' dialog box in Alma. The top screenshot shows the form with empty fields: 'Name*' is an empty text box, 'Description:*' is an empty text area, 'Private' is selected with a radio button, and 'Enabled' is unchecked. The bottom screenshot shows the form filled with the following information: 'Name*' is 'KDNP Title Test', 'Description:*' is 'Test converting title field in generic XML to DC title', 'Shared' is selected with a radio button, and 'Enabled' is checked with a checkbox. Both screenshots include 'Save' and 'Cancel' buttons at the bottom.

- First attempt using documentation at ([https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_\(English\)/100Loading_Records_from_External_Sources_into_Primo_VE/Configuring_Normalization_Rules_for_External_Resources_\(Primo_VE\)#Examples](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/100Loading_Records_from_External_Sources_into_Primo_VE/Configuring_Normalization_Rules_for_External_Resources_(Primo_VE)#Examples))
- Looking at one field (doctitle_t), tried creating rules based on this formula:

<doctitle_t>The Adair County news., 1908-04-08</doctitle_t>

```
rule "move relation to local field 1"
  when
    exist "dc"."relation"
  then
    move "dc"."relation" to "discovery"."local1"
  end
```

- After more searching, I found documentation here (https://knowledge.exlibrisgroup.com/Alma/Community_Knowledge/Generic_XML_Normalization_Rules_and_XPath_Tips) – Knowledge Center article, “Generic XML Normalization Rules and Xpath Trips”; section “Attribute Text as Condition”

Created By: Joan Kolarik
Created on: 2/08/2020

Please send suggestions and improvements to joan.kolarik@weizmann.ac.il

```
rule "rights Closed"  
when  
  exist "//*[local-name()='accessCondition'][@type='info:eu-repo/semantics/closedAccess']"  
then  
  set "info:eu-repo/semantics/closedAccess" in "dc"."rights"  
end
```

KDNP Title Test


```
rule "title"  
when  
  exist "//*[local-name()='doctitle_t']"  
then  
  copy "//*[local-name()='doctitle_t']" to "dc"."title"  
end
```

Save


Save and test using external record

Testing External Discovery Records

Select record type DC Generic XML Marc21

Browse for file  Upload

Select test type Normalization rule Normalization process

Normalization rule *  Run test

Source Record

```
<?xml version = "1.0" encoding = "UTF-8" standalone = "yes" ?>
<collection>
<newspaperissue>
<lccn_link_s>http://chroniclingamerica.loc.gov/lccn/sn86069496/</lccn_link_s>
<aip_t>xt70rx938d1v</aip_t>
<id>xt700000n98</id>
<lccn_t>sn86069496</lccn_t>
<printurl_t>https://archive.org/download/xt700000n98/xt700000n98.pdf</printurl_t>
<thumbnailurl_t>https://www.archive.org/download/xt700000n98/page/n0_w100.jpg</thumbnailurl_t>
<objecturl_t>https://kentuckynewspapers.org/view.php?id=xt700000n98</objecturl_t>
<abbreviation_t>ada</abbreviation_t>
<control_t>ada1908040801</control_t>
<edition_s>01</edition_s>
<collection_s>Adair County news (The)</collection_s>
<title_t>Adair county news (The)</title_t>
<region_s>Pennyrile</region_s>
<dc:title_t>The Adair County news., 1908-04-08</dc:title_t>
```

Normalized Record

dc:title The Adair County news., 1908-04-08

Normalized Record

dc:title The Adair County news., 1908-04-08

Source Record

```

<?xml version = "1.0" encoding = "UTF-8" standalone = "yes" ?>
<collection>
<newspaperissue>
<lccn_link_s>http://chroniclingamerica.loc.gov/lccn/sn86069496/</lccn_link_s>
<aip_t>xt70rx938d1v</aip_t>
<id>xt7000000n98</id>
<lccn_t>sn86069496</lccn_t>
<printurl_t>https://archive.org/download/xt7000000n98/xt7000000n98.pdf</printurl_t>
<thumbnailurl_t>https://www.archive.org/download/xt7000000n98/page/n0_w100.jpg</thumbnailurl_t>
<objecturl_t>https://kentuckynewspapers.org/view.php?id=xt7000000n98</objecturl_t>
<abbreviation_t>ada</abbreviation_t>
<control_t>ada1908040801</control_t>
<edition_s>01</edition_s>
<collection_s>Adair County news (The)</collection_s>
<title_t>Adair county news (The)</title_t>
<region_s>Pennyrile</region_s>
<doctitle_t>The Adair County news., 1908-04-08</doctitle_t>
<pubplace_t>Columbia</pubplace_t>
<publisher_t>Chas. S. Harris</publisher_t>
<county_s>Adair</county_s>
<rights_t>http://creativecommons.org/licenses/by/4.0/</rights_t>
<decade_s>1900</decade_s>
<year_s>1908</year_s>
<digital_date_t>2013</digital_date_t>
<issuedate_t>1908-04-08</issuedate_t>
<month_s>04</month_s>
<day_s>08</day_s>

```

Normalized Record

dcterms:relation	https://www.archive.org/download/xt7000000n98/page/n0_w100.jpg
dcterms:format	Newspaper issue
dc:source	xt7000000n98
dc:source	ada
dc:source	sn86069496
dc:source	ada1908040801
dc:rights	http://creativecommons.org/licenses/by/4.0/
dc:publisher	Chas. S. Harris
dc:coverage	1908-04-08
dc:coverage	Columbia
dc:coverage	Pennyrile
dc:coverage	Adair
dc:coverage	1900
dc:creator	Adair County news (The)
dc:identifier	https://kentuckynewspapers.org/view.php?id=xt7000000n98
dc:title	The Adair County news., 1908-04-08

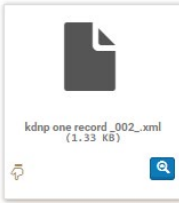
Test Import: Go to Discovery>Loading External Data Sources>Discovery Import Profiles

ID	Status	Profile name	Profile description	Profile Type	Contributed by	
1	Active	ExploreUK	ExploreUK local repository	Discovery	-	...
2	Active	KDNP Newspapers	Import profile for KDNP newspaper issues	Discovery	-	...
3	Active	KDNP Test	-	Discovery	-	...
4	Active	UKnowledge	Harvest from UKnowledge. Not really nee...	Discovery	-	...

- Edit
- View
- History
- Run
- Reload
- Reload and Delete
- Copy
- Delete

Select File

File *



kdnp one record_002_.xml (1.33 KB)

kdnp one record_002_.xml

Label

Test Record

Notes

Add

Start New Import

Cancel Submit

Files List

1 - 1 of 1

File Name	Label	Notes	Source	Size	Date	
kdnp one record (002).xml_1.xml	Test Record	-	-	1 KB	2021/05/19	...

<?xml version="1.0" encoding="UTF-8" standalone="true"?>

<collection>

- <newspaperissue>

<lccn_link_s><http://chroniclingamerica.loc.gov/lccn/sn86069496/></lccn_link_s>

<aip_t>xt70rx938d1v</aip_t>

<id>xt7000000n98</id>

<lccn_t>sn86069496</lccn_t>

<printurl_t><https://archive.org/download/xt7000000n98/xt7000000n98.pdf></printurl_t>

<thumbnailurl_t>https://www.archive.org/download/xt7000000n98/page/n0_w100.jpg</thumbnailurl_t>

<objecturl_t><https://kentuckynewspapers.org/view.php?id=xt7000000n98></objecturl_t>

<abbreviation_t>ada</abbreviation_t>

<control_t>ada1908040801</control_t>

<edition_s>01</edition_s>

<collection_s>Adair County news (The)</collection_s>

<title_t>Adair county news (The)</title_t>

<region_s>Pennyrile</region_s>

<doctitle_t>The Adair County news., 1908-04-08</doctitle_t>

<pubplace_t>Columbia</pubplace_t>

<publisher_t>Chas. S. Harris</publisher_t>

<county_s>Adair</county_s>

<rights_t><http://creativecommons.org/licenses/by/4.0/></rights_t>

<decade_s>1900</decade_s>

<year_s>1908</year_s>

<digital_date_t>2013</digital_date_t>

<issuedate_t>1908-04-08</issuedate_t>

<month_s>04</month_s>

<day_s>08</day_s>

<pages_s>8</pages_s>

<funders_t>National Endowment for the Humanities</funders_t>

<partners_t>Library of Congress</partners_t>

<format_t>Newspaper issue</format_t>

</newspaperissue>

</collection>



NFW5ISSUEIF

The Adair County news., 1908-04-08

Adair County news (The)

[Available Online >](#)

TOP

SEND TO

VIEW ONLINE

DETAILS

TAGS

LINKS

Send to



EXPORT TO EXCEL



EXPORT BIBTEX



EXPORT RIS



EASYBIB



ENDNOTE



REFWORKS



PRINT



CITATION



PERMALINK



EMAIL

View Online

Full text availability

View Resource



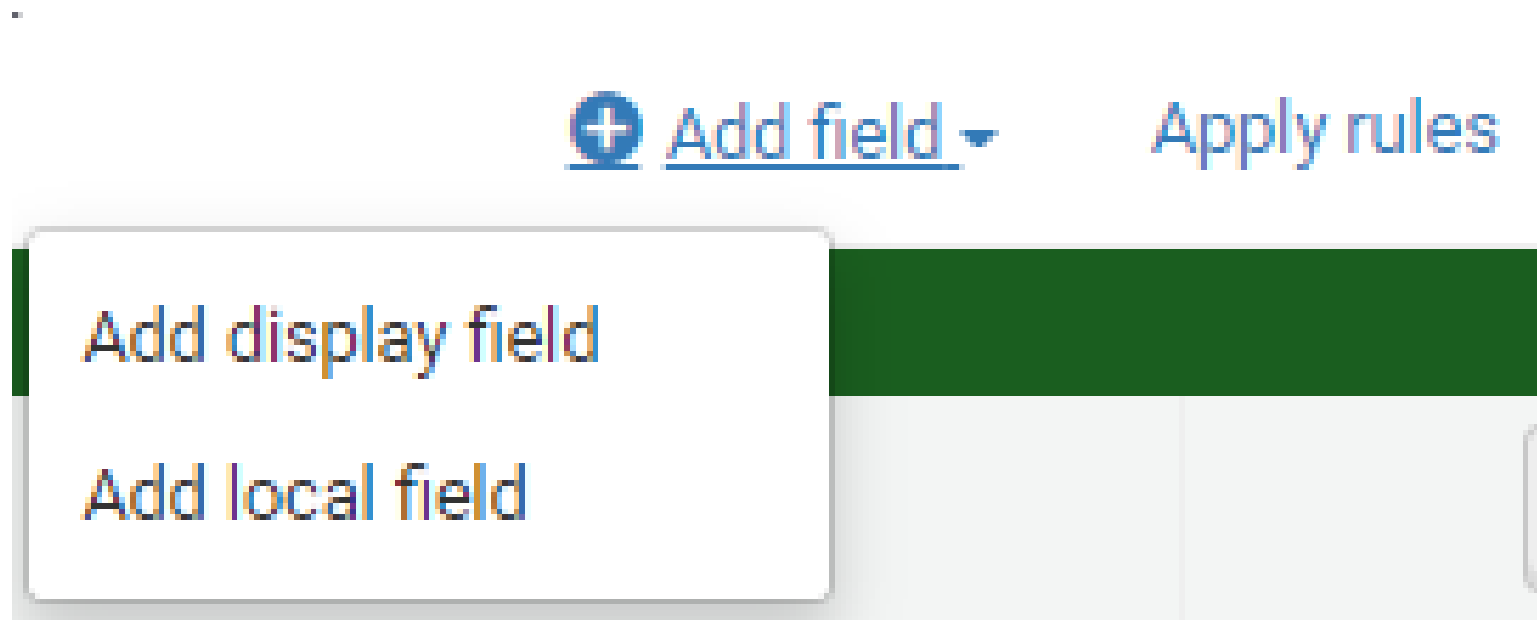
Details

Title	The Adair County news. , 1908-04-08
Creator	Adair County news (The) >
Publisher	Chas. S. Harris
Format	Newspaper issue
Source	KDNPT
Rights	http://creativecommons.org/licenses/by/4.0/

4. Create Local Fields

[https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_\(English\)/050Display_Configuration/040Configuring_Local_Display_and_Search_Fields_for_Primo_VE#Adding_a_Local_Field](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/050Display_Configuration/040Configuring_Local_Display_and_Search_Fields_for_Primo_VE#Adding_a_Local_Field)

- Alma>Discovery>Manage Display and Local Fields>Add Field>Add Local Field



Field to edit * local_field_01

Display label * Rights

Local field details

Enable field for search

Enable field for facet

Use Translations

Use the parallel Local Field 01/50 from the Dublin Core record

Use full text links for indexing

MARC21 Fields

1 - 1 of 1

[+ Add MARC21 Fields](#)

MARC21 Field	
1 009	...

MARC21 Normalization Rules

Normalization rule	Updated by	Updated date	
1 MARC21 normalization rule for display	-	-	...

UNIMARC Fields >

UNIMARC Normalization Rules >

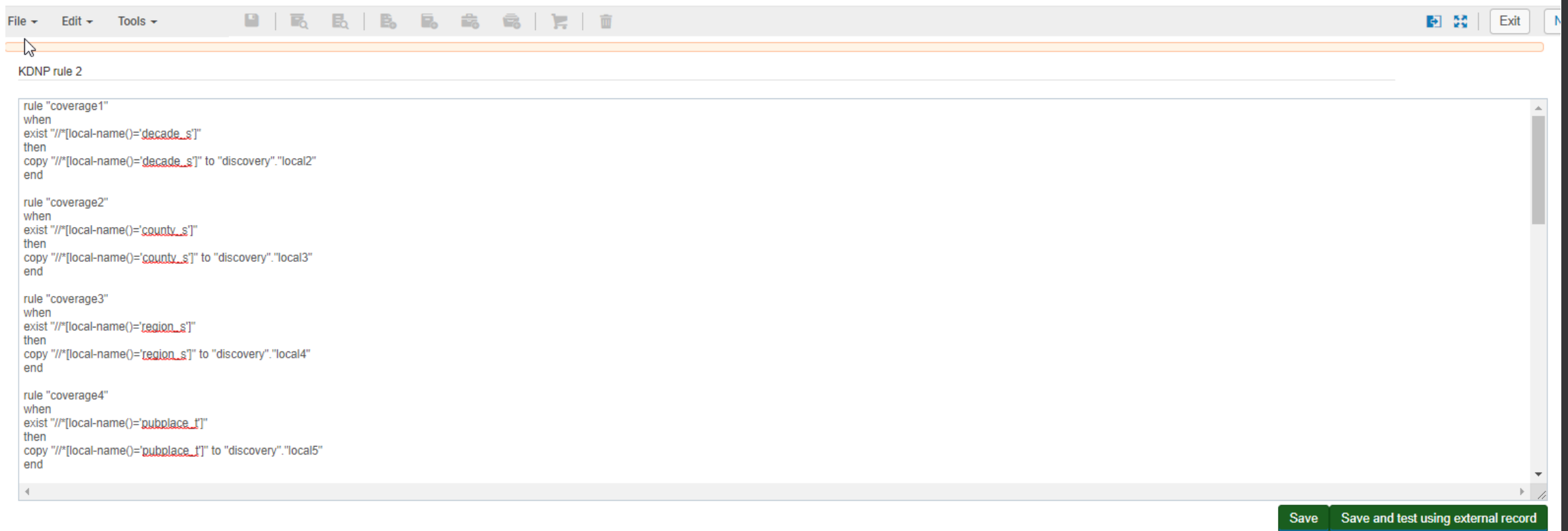
Use the parallel Local Field 01/50 from the Dublin Core record



Use the parallel Local Field 01/50 from the Dublin Core record – For use with Dublin Core records only, select this field if you have defined DC normalization rules that map information to the associated local discovery fields (**discovery.local01** - **discovery.local50**) in the Primo VE DC expanded schema.

	▲ Field	↕ Display label	↕ Updated by	
1	local_field_01	Rights	jmgr235	...
2	local_field_02	Decade	jmgr235	...
3	local_field_03	County	jmgr235	...
4	local_field_04	Region	jmgr235	...
5	local_field_05	Publication Place	jmgr235	...
6	local_field_06	Issue Date	jmgr235	...
7	local_field_07	LCCN	jmgr235	...
8	local_field_08	Control	jmgr235	...
9	local_field_09	Abbreviation	jmgr235	...
10	local_field_10	ID	jmgr235	...
11	local_field_11	Pages	jmgr235	...
12	local_field_12	Edition	jmgr235	...
13	local_field_13	Full-Text KDNP	jmgr235	...

Second Set of Normalization Rules



The image shows a web-based editor interface for KDNP rules. The title bar indicates the file is named "KDNP rule 2". The editor contains four rules, each with a "when" condition and a "then" action. The rules are:

```
rule "coverage1"
when
  exist "//*[local-name()='decade_s']"
then
  copy "//*[local-name()='decade_s']" to "discovery"."local2"
end

rule "coverage2"
when
  exist "//*[local-name()='county_s']"
then
  copy "//*[local-name()='county_s']" to "discovery"."local3"
end

rule "coverage3"
when
  exist "//*[local-name()='region_s']"
then
  copy "//*[local-name()='region_s']" to "discovery"."local4"
end

rule "coverage4"
when
  exist "//*[local-name()='pubplace_!']"
then
  copy "//*[local-name()='pubplace_!']" to "discovery"."local5"
end
```

At the bottom right of the editor, there are two green buttons: "Save" and "Save and test using external record".

Source Record

```

<?xml version = "1.0" encoding = "UTF-8" standalone = "yes" ?>
<collection>
<newspaperissue>
<lcen_link_s>http://chroniclingamerica.loc.gov/lccn/sn86069496/</lcen_link_s>
<aip_t>xt70rx938d1v</aip_t>
<id>xt7000000n98</id>
<lcen_t>sn86069496</lcen_t>
<printurl_t>https://archive.org/download/xt7000000n98/xt7000000n98.pdf</printurl_t>
<thumbnailurl_t>https://www.archive.org/download/xt7000000n98/page/n0_w100.jpg</thumbnail
url_t>
<objecturl_t>https://kentuckynewspapers.org/view.php?id=xt7000000n98</objecturl_t>
<abbreviation_t>ada</abbreviation_t>
<control_t>ada1908040801</control_t>
<edition_s>01</edition_s>
<collection_s>Adair County news (The)</collection_s>
<title_t>Adair county news (The)</title_t>
<region_s>Pennyrile</region_s>
<doctitle_t>The Adair County news., 1908-04-08</doctitle_t>
<pubplace_t>Columbia</pubplace_t>
<publisher_t>Chas. S. Harris</publisher_t>
<county_s>Adair</county_s>
<rights_t>http://creativecommons.org/licenses/by/4.0/</rights_t>
<decade_s>1900</decade_s>
<year_s>1908</year_s>
<digital_date_t>2013</digital_date_t>
<issuedate_t>1908-04-08</issuedate_t>
<month_s>04</month_s>
<day_s>08</day_s>
<pages_s>8</pages_s>
<funders_t>National Endowment for the Humanities</funders_t>

```

Normalized Record

discovery:local1	http://creativecommons.org/licenses/by/4.0/
discovery:local13	https://archive.org/download/xt7000000n98/xt7000000n98.pdf
discovery:local12	01
discovery:local11	8
discovery:local10	xt7000000n98
discovery:local9	ada
discovery:local8	ada1908040801
discovery:local7	sn86069496
discovery:local6	1908-04-08
discovery:local5	Columbia
discovery:local4	Pennyrile
discovery:local3	Adair
discovery:local2	1900

3. Create Normalization Process

Go to Configuration>Discovery>Loading External Data Sources>Normalization Process Task

Click on Add Process

Process List

Description

[+ Add Process](#)

Business Entity : **Discovery BIB Records** ▾



No records were found.

< Process Details

Business Entity *

Type *

< Process Details - General Information

Name *

Description *

Status

Status Date -

[+ Add Tasks](#)

Add Tasks



Name	Description
Generic XML-Normalization	Import generic XML Drool Normalization

Copy

Add Close Add and Close

	Name	Description	
1	Generic XML- Normalization	Import generic XML Drool Normalization	...
2	Generic XML- Normalization	Import generic XML Drool Normalization	...

[Process Details - Task Parameters](#)

1 2 3 4 Cancel Back Save

GENERIC XML DROOL NORMALIZATION

Drools File Key * **kdnf test**

GENERIC XML DROOL NORMALIZATION

Drools File Key * **KDNP rule 2**

Discovery Import Profile (Normalization)

< Import Profile Details

Save Draft

Cancel

Save

Profile Details

Normalization

Delivery

Normalization

Normalize the data using

KDNP



NFW5ISSIIF

The Adair County news., 1908-04-08

Adair County news (The)

[Available Online >](#)



TOP

SEND TO

VIEW ONLINE

DETAILS

TAGS

LINKS

Send to



EXPORT TO EXCEL



EXPORT BIBTEX



EXPORT RIS



EASYBIB



ENDNOTE



REWORKS



PRINT



CITATION



PERMALINK



EMAIL

View Online

Full text availability

[View Resource](#)



Details

Title	The Adair County news., 1908-04-08
Creator	Adair County news (The) >
Publisher	Chas. S. Harris
Format	Newspaper issue
Source	KDNPT
Rights	http://creativecommons.org/licenses/by/4.0/
Decade	1900
Abbreviation	ada
ID	xt7000000n98
LCCN	sn86069496
County	Adair
Region	Pennyryle
Publication Place	Columbia
Issue Date	1908-04-08
Control	ada1908040801
Pages	8
Edition	01

5. Create Local Resource Type

[https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_\(English\)/050Display_Configuration/050Configuring_Local_Resource_Types_for_Primo_VE](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/050Display_Configuration/050Configuring_Local_Resource_Types_for_Primo_VE)

Local Resource Types

[+ Add local Resource Type](#)



No records were found.

Configuration>Discovery>Local Resource Types


Click Add local Resource Type

General

Code *	<input type="text" value="NEWSISSUE"/>	Active	<input checked="" type="checkbox"/>
Display Singular Label *	<input type="text" value="Newspaper Issue"/>	Display Plural Label *	<input type="text" value="Newspaper Issues"/>
RIS type	<input type="text" value="Newspaper"/>	Genre	<input type="text" value="Issue"/>
Metadata format	<input type="text" value="Journal"/>		

Mapping from MARC records

[+ Add Condition](#)


No records were found.

Mapping from Dublin Core

1 - 1 of 1 [+ Add Condition](#)  

Logic	Dublin Core tag	Value in Dublin Core	
1 -	dcterms:format	Newspaper issue	



NEWSISSUE ← **OTB**

The Adair County news., 1908-04-08

Adair County news (The)

[Available Online >](#)



NEWSPAPER ISSUE ← **Customization**

The Adair County news., 1908-04-08

Adair County news (The)

[Available Online >](#)

Discovery Import Profile (Delivery)

< Import Profile Details

Save Draft

Cancel

Save

Profile Details Normalization Delivery

Link to Resource

Template Static URL from source

Dublin Core tag * X ≡ ↻

Link Label ⓘ

Link to Thumbnail

Template Static URL from source

Template

Link to Request

Template Static URL from source

Template

Link Label ⓘ

Linking Parameters ⓘ

Parameter	Source Tag	
1 SourceRecordID	-	...
2 LinkingParameter1	-	...
3 LinkingParameter2	-	...
4 LinkingParameter3	-	...
5 LinkingParameter4	-	...
6 LinkingParameter5	-	...

6. Map Thumbnail Image



Issues:

1. File extension (supported formats - .bmp, .png, .jpg *and so forth*)
2. Normalize XML field
3. Add DC field to Import Profile

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <collection>
  - <newspaperissue>
    <lccn_link_s>http://chroniclingamerica.loc.gov/lccn/sn86069496/</lccn_link_s>
    <aip_t>xt70rx938d1v</aip_t>
    <id>xt7000000n98</id>
    <lccn_t>sn86069496</lccn_t>
    <printurl_t>https://archive.org/download/xt7000000n98/xt7000000n98.pdf</printurl_t>
    <thumbnailurl_t>https://www.archive.org/download/xt7000000n98/page/n0_w150</thumbnailurl_t>
    <objecturl_t>https://kentuckynewspapers.org/view.php?id=xt7000000n98</objecturl_t>
    <abbreviation_t>ada</abbreviation_t>
    <control_t>ada1908040801</control_t>
    <edition_s>01</edition_s>
    <collection_s>Adair County news (The)</collection_s>
    <title_t>Adair county news (The)</title_t>
    <region_s>Pennyrile</region_s>
    <doctitle_t>The Adair County news., 1908-04-08</doctitle_t>
    <pubplace_t>Columbia</pubplace_t>
    <publisher_t>Chas. S. Harris</publisher_t>
    <county_s>Adair</county_s>
    <rights_t>http://creativecommons.org/licenses/by/4.0/</rights_t>
    <decade_s>1900</decade_s>
    <year_s>1908</year_s>
    <digital_date_t>2013</digital_date_t>
    <issuedate_t>1908-04-08</issuedate_t>
    <month_s>04</month_s>
    <day_s>08</day_s>
    <pages_s>8</pages_s>
    <funders_t>National Endowment for the Humanities</funders_t>
    <partners_t>Library of Congress</partners_t>
    <format_t>Newspaper issue</format_t>
  </newspaperissue>
</collection>
```

1.

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <collection>
  - <newspaperissue>
    <lccn_link_s>http://chroniclingamerica.loc.gov/lccn/sn86069496/</lccn_link_s>
    <aip_t>xt70rx938d1v</aip_t>
    <id>xt7000000n98</id>
    <lccn_t>sn86069496</lccn_t>
    <printurl_t>https://archive.org/download/xt7000000n98/xt7000000n98.pdf</printurl_t>
    <thumbnailurl_t>https://www.archive.org/download/xt7000000n98/page/n0_w100.jpg</thumbnailurl_t>
    <objecturl_t>https://kentuckynewspapers.org/view.php?id=xt7000000n98</objecturl_t>
    <abbreviation_t>ada</abbreviation_t>
    <control_t>ada1908040801</control_t>
    <edition_s>01</edition_s>
    <collection_s>Adair County news (The)</collection_s>
    <title_t>Adair county news (The)</title_t>
    <region_s>Pennyrile</region_s>
    <doctitle_t>The Adair County news., 1908-04-08</doctitle_t>
    <pubplace_t>Columbia</pubplace_t>
    <publisher_t>Chas. S. Harris</publisher_t>
    <county_s>Adair</county_s>
    <rights_t>http://creativecommons.org/licenses/by/4.0/</rights_t>
    <decade_s>1900</decade_s>
    <year_s>1908</year_s>
    <digital_date_t>2013</digital_date_t>
    <issuedate_t>1908-04-08</issuedate_t>
    <month_s>04</month_s>
    <day_s>08</day_s>
    <pages_s>8</pages_s>
    <funders_t>National Endowment for the Humanities</funders_t>
    <partners_t>Library of Congress</partners_t>
    <format_t>Newspaper issue</format_t>
  </newspaperissue>
</collection>
```

2.

```
rule "thumbnail"
when
  exist "[local-name()='thumbnailurl_t']"
then
  copy "[local-name()='thumbnailurl_t']" to "dcterms"."relation"
end
```

3.

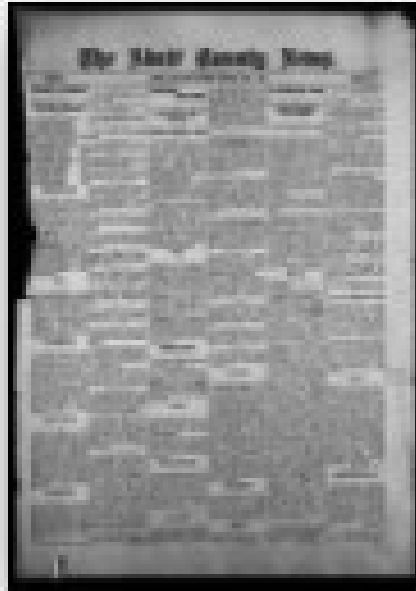
Profile Details Normalization Delivery

Link to Resource

Link to Thumbnail

Template Static URL from source

Dublin Core tag * X ☰ ↺




NEWSPAPER ISSUE

The Adair County news., 1908-04-08

Adair County news (The)

[Available Online >](#)


1



NEWSPAPER ISSUE
The Adair County news., 1908-04-08
 Adair County news (The)
[Available Online](#) >

Brief Record View in Primo VE

NEWSPAPER ISSUE
The Adair County news., 1908-04-08
 Adair County news (The)
[Available Online](#) >



TOP
 SEND TO
 VIEW ONLINE
 DETAILS
 TAGS
 LINKS

Send to

EXPORT TO EXCEL EXPORT BIBTEX EXPORT RIS EASYBIB ENDNOTE REFWORKS PRINT CITATION PERMALINK
 EMAIL

View Online

Full text availability

View at KDNP

Details

Title	The Adair County news., 1908-04-08
Creator	Adair County news (The) >
Publisher	Chas. S. Harris
Format	Newspaper issue
Source	KDNPT
Rights	http://creativecommons.org/licenses/by/4.0/
Decade	1900
Abbreviation	ada
ID	xt7000000n98
LCCN	sn86069496
County	Adair
Region	Pennyrile
Publication Place	Columbia
Issue Date	1908-04-08
Control	ada1908040801
Pages	8
Edition	01

Full Record View in Primo VE

```
<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/" xmlns:discovery="http://purl.org/dc/elements/1.1/">
  <dcterms:relation>https://www.archive.org/download/xt7000000n98/page/n0_w100.jpg</dcterms:relation>
  <dcterms:format>Newspaper issue</dcterms:format>
  <dc:source>xt7000000n98</dc:source>
  <dc:source>ada</dc:source>
  <dc:source>sn86069496</dc:source>
  <dc:source>ada1908040801</dc:source>
  <dc:rights>http://creativecommons.org/licenses/by/4.0/</dc:rights>
  <dc:publisher>Chas. S. Harris</dc:publisher>
  <dc:coverage>1908-04-08</dc:coverage>
  <dc:coverage>Columbia</dc:coverage>
  <dc:coverage>Pennyrile</dc:coverage>
  <dc:coverage>Adair</dc:coverage>
  <dc:coverage>1900</dc:coverage>
  <dc:creator>Adair County news (The)</dc:creator>
  <dc:identifier>https://kentuckynewspapers.org/view.php?id=xt7000000n98</dc:identifier>
  <dc:title>The Adair County news., 1908-04-08</dc:title>
  <discovery:local1>http://creativecommons.org/licenses/by/4.0/</discovery:local1>
  <discovery:local13>https://archive.org/download/xt7000000n98/xt7000000n98.pdf</discovery:local13>
  <discovery:local12>01</discovery:local12>
  <discovery:local11>8</discovery:local11>
  <discovery:local10>xt7000000n98</discovery:local10>
  <discovery:local9>ada</discovery:local9>
  <discovery:local8>ada1908040801</discovery:local8>
  <discovery:local7>sn86069496</discovery:local7>
  <discovery:local6>1908-04-08</discovery:local6>
  <discovery:local5>Columbia</discovery:local5>
  <discovery:local4>Pennyrile</discovery:local4>
  <discovery:local3>Adair</discovery:local3>
  <discovery:local2>1900</discovery:local2>
</record>
```

One final obstacle - Harvesting Full Text for Search

- Goal is to be able to search keywords/phrases to locate an article/issue from the collection
- Hit a snag with this – still investigating

[https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_\(English\)/050Display_Configuration/040Configuring_Local_Display_and_Search_Fields_for_Primo_VE#Adding_a_Local_Field](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/050Display_Configuration/040Configuring_Local_Display_and_Search_Fields_for_Primo_VE#Adding_a_Local_Field)

Use full text links for indexing – Select this option if you want to index the externally-held full text file (which can be of type PDF, TXT, or HTML) to which this field links in the bibliographic record. During the indexing of the local fields, Primo VE will perform the following actions on each of the full text records:

- Remove stop words based on the language.
- Remove HTML tags.
- Index up to 10,000 terms.

<printurl_t>https://archive.org/download/xt7000000n98/xt7000000n98.pdf</printurl_t>

```
rule "fulltext"  
when  
  exist "//*[local-name()='printurl_t']"  
then  
  copy "//*[local-name()='printurl_t']" to "discovery"."local13"  
end
```

13	local_field_13	Full-Text KDNP
----	----------------	----------------

Field to edit * Display label *

Local field details

Enable field for search Use the parallel Local Field 01/50 from the Dublin Core record

Enable field for facet Use full text links for indexing

Use Translations

MARC21 Normalization Rules

Normalization rule	Updated by	Updated date	
1 MARC21 normalization rule for display	-	-	...
2 MARC21 normalization rule for search and facet	-	-	...

UNIMARC Normalization Rules



Questions?
Thank you!