

# Newspaper Metadata Manual

The weekly summer edition  
www.ntdaily.com  
Thursday • June 10, 2004

**North Texas Daily**  
Student Newspaper at the University of North Texas



**Christine Stanley**  
Assistant Managing Editor

While the illustrious poodle "Legally Blonde" Lily was being primped for the spokesdog pageant, a retriever drank from bottled water in front of Beth Marie's. Basset Hounds sauntered across the courthouse lawn with massive Great Danes and Puggys, the celebrity Pekinese, waited in line for his "Pet Star" audition.

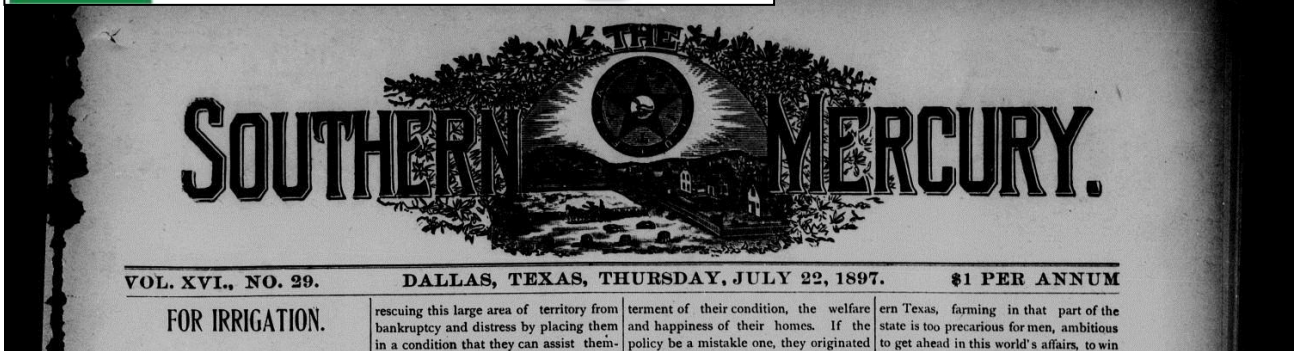
This was only a fraction of the spectacle that made up the 11th annual Dog Days of Summer celebration, held Saturday, June 5. The downtown square was closed to traffic from 9 a.m. to 4 p.m. in honor of area canines and their owners. Christine Gossett, Denton Main Street Marketing Specialist, estimates that about 9,000

See DOGS, Page 20

BECKM-TURNER/NT DAILY  
A.J., a one-year-old boxer poses for a picture at the Dog Days of Summer Saturday morning.

It's a Dog's life

```
<?xml version="1.0" encoding="UTF-8" ?>
<metadata>
<title qualifier="officialtitle">The Democrat (McKinne
<title qualifier="serialtitle">The Democrat</title>
<creator qualifier="edt">
<type>per</type>
<name>Smith, J. Frank</name>
</creator>
<contributor>
<info></info>
<type></type>
<name></name>
</contributor>
<publisher>
<location>McKinney, Texas</location>
<name>Wilson & Smith</name>
</publisher>
<date qualifier=""></date>
<language>eng</language>
<description qualifier="content">Weekly newspaper from
<description qualifier="physical">pages : ill. ; page
<subject qualifier="UNTL-BS">Business, Economics and F
<subject qualifier="UNTL-BS">Business, Economics and F
<subject qualifier="UNTL-BS">Business, Economics and F
<subject qualifier="UNTL-BS">Places - United States -
<subject qualifier="LCSH">Collin County (Tex.) -- News
<subject qualifier="LCSH">McKinney (Tex.) -- Newspaper
<subject qualifier="LCSH">McKinney (Tex.) -- Periodics
<primarySource>1</primarySource>
<coverage qualifier="placeName">United States - Texas
<coverage qualifier="timePeriod">new-sou</coverage>
<source qualifier=""></source>
<citation qualifier=""></citation>
<relation qualifier=""></relation>
<collection>TDNP</collection>
<collection>MKDGTZ</collection>
<institution>CCGS</institution>
<rights qualifier=""></rights>
<resourceType>text_newspaper</resourceType>
<format>text</format>
<identifier qualifier="LCCN">sn86089835</identifier>
<identifier qualifier="OCLC">14564037</identifier>
```



**THE SOUTHERN MERCURY.**

VOL. XVI., NO. 29. DALLAS, TEXAS, THURSDAY, JULY 22, 1897. \$1 PER ANNUM

FOR IRRIGATION. rescuing this large area of territory from bankruptcy and distress by placing them in a condition that they can assist them-  
 ment of their condition, the welfare and happiness of their homes. If the policy be a mistake one, they originated  
 ern Texas, farming in that part of the state is too precarious for men, ambitious to get ahead in this world's affairs, to win

THE PORTAL TO TEXAS HISTORY

University of North Texas Libraries: Digital Newspaper Unit

Texas Digital Newspaper Program



# Table of Contents

Introduction to Newspaper Metadata.....	1
Revision History .....	1
Chapter 1: Metadata Production .....	3
Microfilm Evaluation.....	3
Microfilm Reader.....	3
Collation Spreadsheet .....	4
Reel Metadata .....	5
Collation Data .....	7
Issue Metadata .....	8
Separate Issues.....	8
Create Reports.....	10
Chapter 2: Metadata Quality Control .....	11
Student Metadata QC.....	11
Staff Metadata QC .....	13
AutoHotkey Tools .....	13
Chapter 3: Super_Metadata .....	15
Super_Metadata Templates .....	15
New Record Creator .....	16
Chapter 4: Metadata Post-Processing .....	19
Portal to Texas History Dashboard .....	19
Selenium IDE Automation.....	19
Conclusion.....	21
References .....	21
Appendix: Newspaper Metadata Glossary .....	23



## Introduction to Newspaper Metadata

This document describes the process that the University of North Texas Libraries' Digital Newspaper Unit (DNU) uses to create metadata for the scanned newspapers published on the Portal to Texas History as part of the Texas Digital Newspaper Program (TDNP). High quality metadata helps users find what they want more quickly. Over time, accurate descriptive metadata fosters trust among the Portal's user community.

The DNU creates metadata in four stages:

1. We gather microfilm collation data, divide digital files into issue folders, and create metadata for each issue.
2. We check the metadata in a two-step process.
3. We create metadata templates that apply descriptive metadata to groups of similar issues during the Portal to Texas History upload process.
4. We use Web automation software to correct metadata problems discovered in the digital newspapers published on the Portal to Texas History.

Each chapter covers one of the stages outlined above with detailed instructions for the workflows involved in producing and editing digital newspaper metadata. The References section at the end contains URLs for useful resources on the Web. The Appendix consists of a glossary which defines commonly used technical vocabulary.

## Revision History

Version 0.1: Created by Andrew Weidner. Submitted as the final project for UNT's TECM 2700: Technical Writing. December 2012.

Version 0.5: Edited by Andrew Weidner. Added new text and graphics. Submitted to the Digital Newspaper Unit staff for review. January 2013.

Version 1.0: Edited by Andrew Weidner. Incorporated staff comments. Special thanks to Ana Krahmer, Ann Howington, Skye Limon, Olivia Yeatts and Karen Quinn for their helpful suggestions. March 2013.

*Cover images:*

North Texas Daily, June 10, 2004. Retrieved from: <http://texashistory.unt.edu/ark:/67531/metaph145141/>

Southern Mercury, July 22, 1897. Retrieved from: <http://texashistory.unt.edu/ark:/67531/metaph185719/>



# Chapter 1: Metadata Production

The DNU uses a two-step process to produce basic descriptive metadata for its digital newspapers. The first step gathers general information about microfilm reels with a microfilm reader and a spreadsheet. The second step creates basic metadata for each issue and prepares the digital objects for further processing and description.

## Microfilm Evaluation

We digitize most of our newspapers from 35 mm microfilm negatives with a high speed microfilm scanner. Since microfilm quality varies greatly, we view each reel with a microfilm reader and gather data about any problems that we find. We collate the data in spreadsheets as the first step in digital newspaper metadata production.

### Microfilm Reader

Before viewing film in the microfilm reader, turn on the reader lamp to check for dirt on the glass plates. If there are visible spots projected onto the screen, use the microfilm reader cleaning kit to clean the glass plates. After making sure that the glass plates are clean, load a reel in the microfilm reader with the following steps:

1. Remove the microfilm reader's dust cover and take the reel out of its box.
2. Pull the reader assembly out until the glass plate opens completely.
3. Place the reel on the left hand spindle. The film's loose end should trail over the reel's top edge from left to right.
4. Thread the film under the glass plate and through the six rollers in the following order: over, under, over, over, under, over.
5. Insert the film in the slot on the empty reel.
6. Gently turn the handle on the empty reel clockwise until the film wraps securely around the reel.
7. Push the reader assembly in until the glass plate closes completely, and turn on the reader's lamp with the power switch.

After completing the evaluation for a reel, unload the reel with the following steps:

1. Turn off the reader lamp with the power switch and move the reel handle to the left hand reel.
2. Pull the reader assembly out until the glass plate opens completely and rewind the film. Remove the reel from the left hand spindle and place it in its box.
3. Push in the reader assembly until the glass plate closes completely and replace the dust cover.

## Collation Spreadsheet

We gather collation data on a custom Microsoft Excel spreadsheet designed specifically for the Texas Digital Newspaper Program's descriptive metadata system. Follow the steps below to prepare a new microfilm collation spreadsheet:

1. Open the following folder in Windows Explorer: Q:/TDNP/Collation Sheets/
2. Copy the following file: Collation\_template\_TDNP.xls
3. Paste Collation\_template\_TDNP.xls in the appropriate project folder, for example: Q:/TDNP/Collation Sheets/Sweetwater
4. Rename Collation\_template\_TDNP.xls according to the UNT control number label on the microfilm reel box (e.g., UNT\_1151.xls). If the box does not have a control number label, remove the next control number sticker from the control number label sheet and place it on one of the box's long edges. The template for printing new labels is available in the following directory: Q:\TDNP\Film evaluation sheets\BOX LABELS.doc
5. Open the spreadsheet file in Microsoft Excel.

The screenshot shows the Microsoft Excel interface with the 'Reel Information' tab selected. The spreadsheet contains the following data:

	A	B	C	D	E	F	G	
1	Roll Information:	Sweetwater						
2	Reel Number/Bar Code:	UNT_1151						
3	Award year:							
4	Project Name:							
5								
	Title /Location	LCCN	OCLC	Start Date YYYY-MM-DD	End Date YYYY- MM-DD	Era	Daily or Weekly?	Loos
6								
7	Sweetwater Reporter	sn86089515	14396347	1950-01-01	1950-03-31	mod-tim	Daily	Loos
8								
9								
10								
11								
12								
13								

Figure 1.1: Reel Information tab of the TDNP collation spreadsheet.



## Reel Metadata

Reel metadata describes the general characteristics of the filmed newspapers and provides data that helps the microfilm scanning technician produce high quality scans. We gather the following data types for each reel and enter them in the collation spreadsheet tab labeled Reel Information (Figure 1.1):

- **Reel Number:** Input the UNT control number from the side of the microfilm reel box. This number helps the scanning technician keep track of the reels.
- **Title/Location:** Use the title listed in the US Newspaper Directory if there is a record for it in the directory. If the newspaper does not have a directory record, check for an OCLC number in the WorldCat database on the UNT Library Web site. Otherwise, use the title printed on the front page of the newspaper. If the city where the newspaper was published is not reflected in the title, include the location in parentheses. The US Newspaper Directory is available at the following URL: <http://chroniclingamerica.loc.gov/search/titles/>
- **LCCN:** Copy the Library of Congress Control Number (LCCN) from the US Newspaper Directory record.
- **OCLC:** Copy the OCLC number from the US Newspaper Directory record.
- **Dates:** Input the Start Date and End Date, as listed on the microfilm reel's box, in the appropriate spreadsheet fields. Enter dates in the following format: YYYY-MM-DD.
- **Era:** Choose the correct time period from the UNT Libraries controlled vocabulary. You may apply more than one if the time periods overlap:  
<http://digital2.library.unt.edu/vocabularies/coverage-eras/>
- **Daily/Weekly:** Determine whether the newspaper was published on a daily or weekly schedule. Consider five or more issues per week as a daily paper. Other options include: Bi-Weekly (every two weeks), Semi-Weekly (twice per week), Tri-Weekly (three times per week), and Monthly.
- **Loose/Bound:** Look at the images on the microfilm to determine whether the newspaper was filmed as loose sheets or whether it was bound in a volume. Most newspapers are filmed as loose sheets, and two pages may be joined at the fold. Bound newspaper images usually have a large dark area in between the joined pages at the spine of the bound volume. The text may also curve or become out of focus toward the center of the bound pages. Loose sheets appear flat.
- **Pages per Issue:** Determine the most common number of pages per issue (i.e., the mode) after completing the reel collation. This number need not be exact. Simply make an estimate based on the page counts recorded in the spreadsheet calendars.
- **Pages per Reel:** Record the total number of pages on the reel. With a calculator, add up the number of pages for each year as indicated in the upper right corner of each year tab.

- **Position:** Determine the page orientation (Figure 1.2). Most newspapers are filmed in the 2B position: upright with two joined pages in a single frame.



Figure 1.2: Microfilm page position codes. (RLG Guidelines, 2003)

- **Reduction Ratio:** The reduction ratio indicates the level of magnification required to reproduce a full size page image (e.g., 20x = magnify 20 times to view full size). The scanning technician uses this number to calibrate the high speed microfilm scanner for each reel. Microfilmmers often include a target at the beginning of a reel that indicates the reduction ratio. If no reduction ratio is indicated, convert the Actual Size of the page from inches to millimeters and divide that number by the Size on Film.
- **Size on Film:** Measure the size of a page image on the film in millimeters with a ruler and a lightbox.
- **Actual Size:** Determine the size of the original newspaper page in inches. If the page size is absent from the 300 field of the US Newspaper Directory MARC record, multiply both the height and width of the Size on Film by the Reduction Ratio to determine the page dimensions in millimeters and then convert to inches. If the reduction ratio is unknown, consult the *Ayer's Guide* on The Portal to Texas History: [http://digital.library.unt.edu/search/?q=%22N.+W.+Ayer+%26+Son%22&sort=date\\_a&t=dc\\_creator](http://digital.library.unt.edu/search/?q=%22N.+W.+Ayer+%26+Son%22&sort=date_a&t=dc_creator)

## Collation Data

Collation data indicates problems with the microfilm that must be corrected before we publish a digital newspaper online. Detailed collation data helps us create digital objects that accurately represent the original newspaper issues. Follow the steps below to gather collation data with the microfilm reader and a collation spreadsheet:

1. Create a new tab for each year represented on the film with the Insert Worksheet button located at the bottom of the spreadsheet (or Shift+F11).
2. Double click each tab label to enter a year.
3. Copy all of the data in the Calendar Template tab: Ctrl+A (possibly twice to copy the entire sheet) then Ctrl+C. Paste the Calendar Template data into each year tab with Ctrl+V.
4. Enter the Title, Reel Number and Year in the top row of each year tab.
5. Slowly advance the microfilm to the front page of the first issue.
6. Select the year tab that matches the date of the first issue.
7. Slowly advance the microfilm, count the number of pages in the issue, and enter the number of pages in the calendar (Figure 1.3).

12	<b>February</b>	1	2	3	4	5	6	7
13	294	8	16	8		16	8	8
14		8	9	10	11	12	13	14
15		8	16	8		28	8	8
16		15	16	17	18	19	20	21
17		8	20	8		24	8	8
18		22	23	24	25	26	27	28
19		8	16	8		24	8	14
20		29						
21								

Figure 1.3: Calendar spreadsheet example.

8. Record any problems with an issue in the Notes tab. Highlight the corresponding date field by setting the background color to yellow. Common problems include:
  - Duplicate pages: List the duplicate page number(s).
  - Missing pages: List the missing page number(s).
  - Incorrect date: Record the number of pages under the correct date in the calendar and indicate the incorrect date printed on the issue in Notes.
9. Repeat steps 7 and 8 until the the end of the reel. Save and close the spreadsheet file.

## Issue Metadata

After the scanning technician has scanned the reel and processed the images, the image files will be located in a folder named according to the reel's UNT control number. This section describes the process for creating metadata for each newspaper issue on a reel.

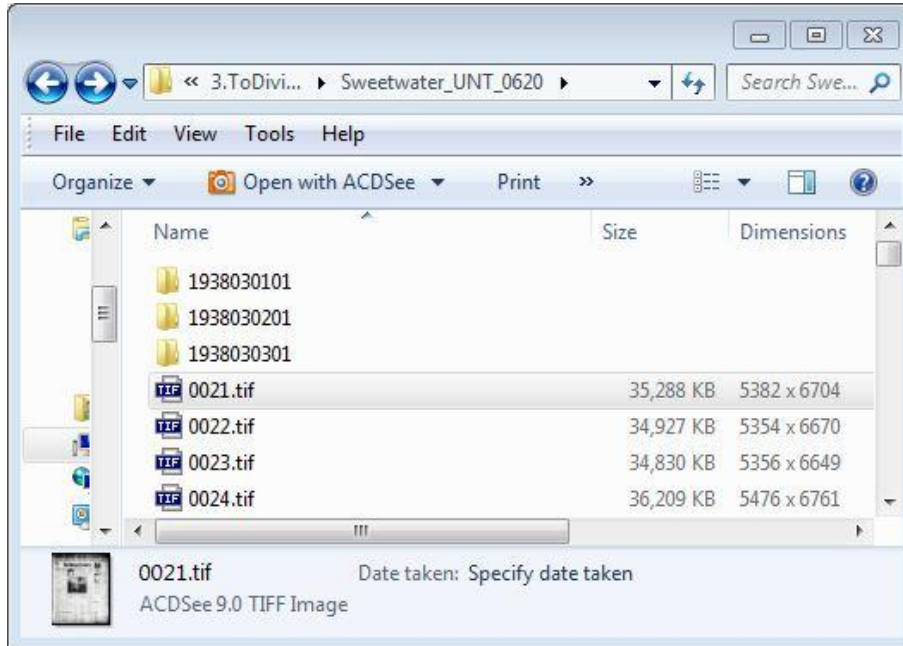


Figure 1.4: Issue separation in progress.

## Separate Issues

The issue separation process creates the date, volume number, and issue number metadata for each issue and prepares the digital object for file replication and optical character recognition processing. Follow the steps below to separate issues:

1. Open the reel folder in Windows Explorer and open the first image file in ACDSee. Create a new folder in the reel folder named with the newspaper title.
2. Note the issue date of the first image and create a new folder named with that date in the following format: YYYYMMDDII. See Figure 1.4 above for an example.
  - Y = Year
  - M = Month
  - D = Day
  - I = Issue Number. Start with 01 for the first issue on a specific date. Use the next number in the sequence (e.g. 02, 03) for any subsequent issues published on the same date.
3. Open the new folder and create a text file called `metadata.txt`.

- Open the `metadata.txt` file and create three lines as shown in Figure 1.5.

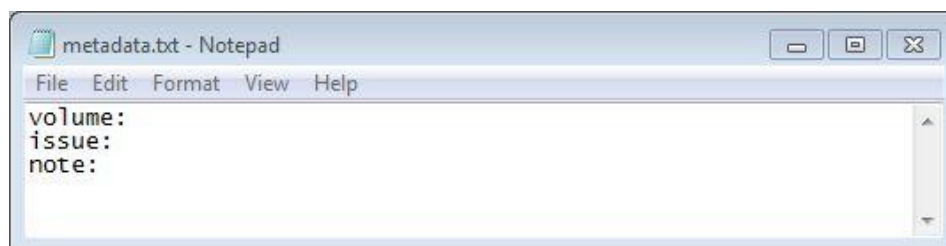


Figure 1.5: Example `metadata.txt` file.

- Enter the volume and issue number in the `metadata.txt` file (Figure 1.6).

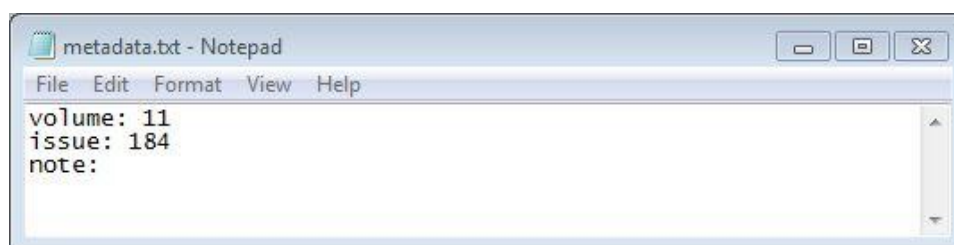


Figure 1.6: Volume and issue numbers entered in the `metadata.txt` file.

- If the date printed on the issue's front page is incorrect, include that information in the note (Figure 1.7). Rename the issue folder to reflect the correct date.



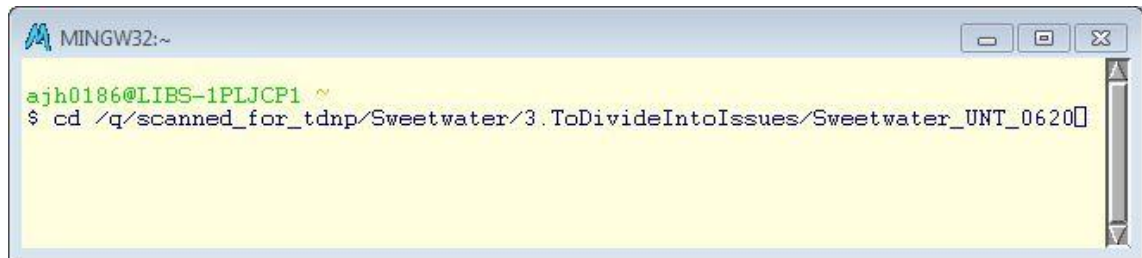
Figure 1.7: Include a note in the `metadata.txt` file for an incorrect issue date.

- Save the `metadata.txt` file and close it. In ACDSee, view each page in the issue with the Page Down key, and remember the file number of the last image in the issue.
- In Windows Explorer, move all images for the issue into its issue folder and move the issue folder into the title folder.
- Repeat steps 2 through 8 until all the image files have been separated into issue folders.
- Label the reel as `Completed` on the Microfilm Scanning Queue wiki page: [http://digitalprojects.library.unt.edu/projects/index.php/Microfilm\\_Scanning\\_Queue](http://digitalprojects.library.unt.edu/projects/index.php/Microfilm_Scanning_Queue)

## Create Reports

After separating all of the issues on a reel, create a report for use during the quality control process. Reports gather information for every issue in a reel folder and display that information in a table. Follow the steps below to create a report:

1. Open an MSYS terminal from the Start menu in Windows.
2. Navigate to the reel folder with the `cd` (change directory) command as shown in Figure 1.8. Press the Enter key to enter the directory. If necessary, use the `ls` (list) command to view the contents of a directory.



```

MINGW32:~
ajh0186@LIBS-1PLJCP1 ~
$ cd /q/scanned_for_tdnf/Sweetwater/3.ToDivideIntoIssues/Sweetwater_UNT_0620

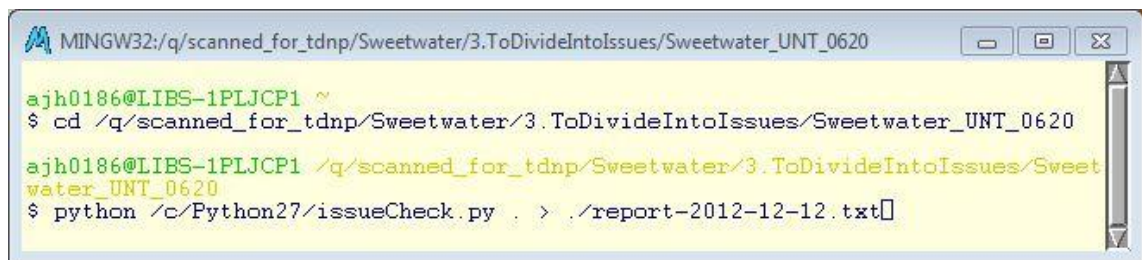
```

Figure 1.8: Navigate to the reel directory in MSYS with the `cd` command.

3. Run the Python report script with the following command:
4. If the script runs successfully, output the report to a text file in the reel folder with the following command (Figure 1.9):

```
python /c/Python27/issueCheck.py .
```

```
python /c/Python27/issueCheck.py . > ./report-yyyy-mm-dd.txt
```



```

MINGW32:/q/scanned_for_tdnf/Sweetwater/3.ToDivideIntoIssues/Sweetwater_UNT_0620
ajh0186@LIBS-1PLJCP1 ~
$ cd /q/scanned_for_tdnf/Sweetwater/3.ToDivideIntoIssues/Sweetwater_UNT_0620
ajh0186@LIBS-1PLJCP1 /q/scanned_for_tdnf/Sweetwater/3.ToDivideIntoIssues/Sweetwater_UNT_0620
$ python /c/Python27/issueCheck.py . > ./report-2012-12-12.txt

```

Figure 1.9: Output the report to a text file with MSYS.

5. Move the reel folder to the `4.1.ToQCMetadata-Students` folder.
6. For subsequent reports, use the Up and Down arrow keys to access previously typed commands in MSYS. This technique also works after exiting and restarting MSYS, so you need only type a long command once and edit it later as necessary.

## Chapter 2: Metadata Quality Control

The metadata quality control (QC) process ensures accurate issue metadata and creates additional descriptive metadata that adds value to the digital newspaper issues. We divide the process into two steps. In the first step, student employees check the issue metadata, make corrections, and create additional descriptive metadata. In the second step, the DNU full-time staff checks the student employees' work.

### Student Metadata QC

The DNU's student employees perform the first stage of metadata quality control. Student employees use the reports to identify and correct common problems with the printed newspapers and/or microfilm. The reports produced at the end of the metadata creation stage provide conveniently formatted tables that allow student employees to compare the metadata for the issues on a reel (Figure 2.1). Follow the steps below to identify and correct common metadata problems:

1. Open a report in a reel folder on the Q: drive.
2. Look for anomalies in the volume number sequence:
  - If the volume number recorded in the metadata is incorrect, input the correct volume number in the `metadata.txt` file.
  - If an incorrect volume number printed on the issue's front page is an isolated or limited occurrence, include the incorrect number in a `metadata.txt` note:
 

note: Incorrect volume number printed on front page: Vol. 11.
  - If an incorrect volume number printed on the issue's front page is continued on subsequent issues, include the following note in the `metadata.txt` file:
 

note: volume number out of sequence.
3. Look for anomalies in the issue number sequence:
  - If the issue number recorded in the metadata is incorrect, input the correct issue number in the issue's `metadata.txt` file.
  - If the incorrect issue number printed on the issue's front page is an isolated or limited occurrence, include the incorrect number in a `metadata.txt` note:
 

note: Incorrect issue number printed on front page: No. 189.
  - If the incorrect issue number printed on the issue's front page begins a new issue number sequence, include the following note in the `metadata.txt` file:
 

note: Issue number out of sequence.

4. Check the day of the week for each issue. If a day seems out of place, check the day and date printed on the front page:
  - If the issue folder's date is incorrect, rename the issue folder.
  - If the date printed on the front page is incorrect, include the incorrect date in a `metadata.txt` note:
 

note: Incorrect date printed on front page: November 5, 1892.
  - If the day printed on the front page is incorrect, include the incorrect day in a `metadata.txt` note:
 

note: Incorrect day printed on front page: Friday, November 5, 1892.
5. Compare the number of images for each issue. If an issue has an odd number of pages, or seems to have more or fewer images than it should, open the issue folder and view the images:
  - If pages are missing from the issue, include the missing page number(s) in a `metadata.txt` note:
 

note: Missing pages 3-6.
  - Delete any duplicate page files. **Double check that a file is actually a duplicate page before removal.** If in doubt, ask a full-time staff member.
6. Open each issue folder and view each page in ACDSee. Create notes for special edition and/or section labels as printed on the page. For example:
  - note: Armistice Day Edition.
  - note: Includes comic section.
7. Create a new report and check your work.
8. Move the reel folder to the `4.2.ToQCMetadata-Fulltime` folder.

```

report-2012-12-11.txt - Notepad
File Edit Format View Help
1938-08-15 Vol. 41 No. 114 Ed. 1 Pgs. 8 Monday xxxxxxxx Incorrect issue number printed on front page: No. 111.
1938-08-16 Vol. 41 No. 115 Ed. 1 Pgs. 6 Tuesday xxxxxxxx Incorrect issue number printed on front page: No. 112.
1938-08-17 Vol. 41 No. 116 Ed. 1 Pgs. 6 wednesday xxxxxxxx Incorrect issue number printed on front page: No. 113.
1938-08-18 Vol. 41 No. 117 Ed. 1 Pgs. 8 Thursday xxxxxxxx Incorrect issue number printed on front page: No. 113.
1938-08-19 Vol. 41 No. 118 Ed. 1 Pgs. 8 Friday xxxxxxxx Incorrect issue number printed on front page: No. 113.
1938-08-21 Vol. 41 No. 119 Ed. 1 Pgs. 20 Sunday xxxxxxxxxxxxxxxxxxxxxxxx Includes comic section.
1938-08-22 Vol. 41 No. 120 Ed. 1 Pgs. 6 Monday xxxxxxxx
1938-08-23 Vol. 41 No. 121 Ed. 1 Pgs. 6 Tuesday xxxxxxxx Incorrect issue number printed on front page: No. 120.
1938-08-24 Vol. 41 No. 122 Ed. 1 Pgs. 6 wednesday xxxxxxxx
1938-08-25 Vol. 41 No. 122 Ed. 1 Pgs. 8 Thursday xxxxxxxx Issue number out of sequence.
1938-08-26 Vol. 41 No. 123 Ed. 1 Pgs. 10 Friday xxxxxxxxxx
1938-08-28 Vol. 41 No. 124 Ed. 1 Pgs. 18 Sunday xxxxxxxxxxxxxxxxxxxxxxxx Includes comic section.
1938-08-29 Vol. 41 No. 125 Ed. 1 Pgs. 6 Monday xxxxxxxx
1938-08-30 Vol. 41 No. 125 Ed. 1 Pgs. 6 Tuesday xxxxxxxx Issue number out of sequence.
1938-08-31 Vol. 41 No. 126 Ed. 1 Pgs. 8 wednesday xxxxxxxxxx
1938-09-01 Vol. 41 No. 127 Ed. 1 Pgs. 8 Thursday xxxxxxxxxx
1938-09-02 Vol. 41 No. 128 Ed. 1 Pgs. 8 Friday xxxxxxxxxx
1938-09-06 Vol. 41 No. 129 Ed. 1 Pgs. 6 Tuesday xxxxxxxx
  
```

Figure 2.1: Example report with metadata corrections and descriptive notes.



## Staff Metadata QC

The DNU's full-time staff performs the second stage of metadata quality control. The full-time staff checks the student employees' work and makes changes to the metadata as necessary. Follow the steps below to perform the Staff Metadata QC:

1. Open a report in a reel folder.
2. Check the report for obvious errors in the volume and issue number sequence. Correct errors as described in steps 2 and 3 of the Student Metadata QC section above.
3. Check the report for obvious errors in the day, date, and number of image files. Correct errors as described in steps 4 and 5 of the Student Metadata QC section above.
4. Check the report for typographical errors and content accuracy in the notes. If necessary, make corrections to the `metadata.txt` files.
5. Spot check approximately 10% of the issues in the reel as described in step 6 of the Student Metadata QC section above.
6. Move the reel folder to the 5.ToOCR folder.

## AutoHotkey Tools

The DNU uses AutoHotkey software applications to automate various aspects of the newspaper metadata production process. These applications are available on the Q: drive in the Software folder. In general, the DNU's AutoHotkey applications save time and improve metadata quality. For example, the `NewspaperNotes.exe` tool (Figure 2.2) automatically types commonly used notes according to the standard formulation listed on the Newspaper Notes wiki:

[http://digitalprojects.library.unt.edu/projects/index.php/Newspaper\\_Notes](http://digitalprojects.library.unt.edu/projects/index.php/Newspaper_Notes)

The next page describes the `TDNP_Microfilm.exe` and `TDNP_Metadata.exe` AutoHotkey applications. Together with `NewspaperNotes.exe`, these applications streamline the workflows outlined in the first two chapters.

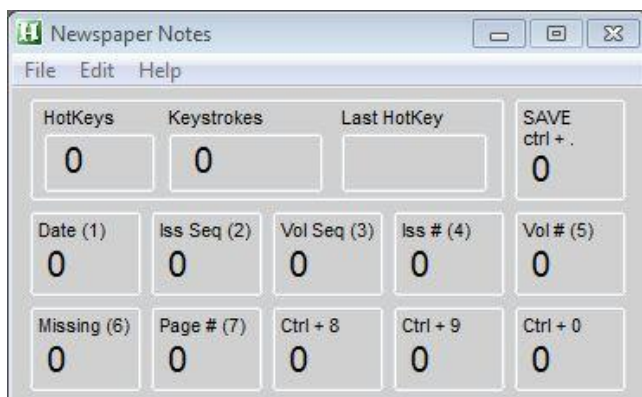


Figure 2.2: The `NewspaperNotes.exe` tool.

## TDNP\_Microfilm.exe

The TDNP\_Metadata.exe application (Figure 2.3) is designed for the microfilm evaluation tasks described in Chapter 1. It bundles together three hotkeys for microfilm evaluation with the TDNP collation spreadsheet.

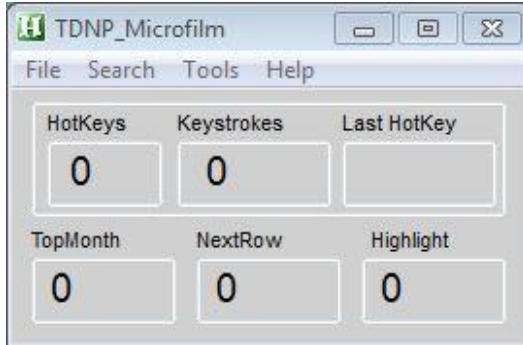


Figure 2.3: The TDNP\_Microfilm.exe window.

The File menu has links to the Collation Sheets and scanned\_for\_tdnf folders on the Q: drive. The Search menu provides search functions for the U.S. Newspaper Directory, the Portal to Texas History, and the Gateway to Oklahoma History. The Tools menu contains tools to assist with measurement conversions and reduction ratio and page size calculations.

## TDNP\_Metadata.exe

The TDNP\_Metadata.exe application (Figure 2.4) is designed for the issue separation tasks described in Chapter 1 and the quality control tasks described in Chapter 2. It contains three quality control hotkeys that navigate between folders and open files for viewing (Open, Next, & Previous). A fourth hotkey opens the metadata.txt file for editing (Edit Meta). It also has a File menu option to easily create the reports described on page 10.

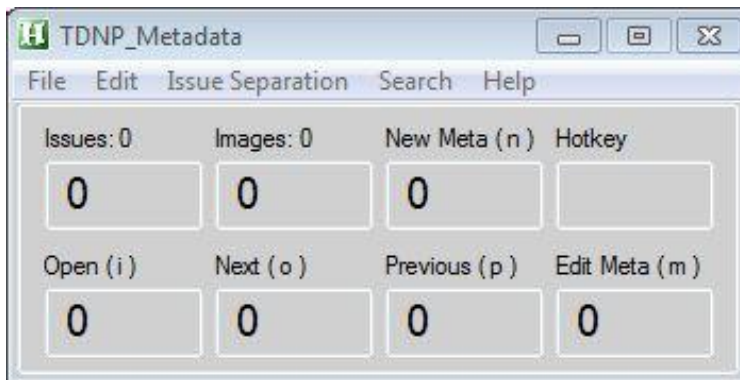


Figure 2.4: The TDNP\_Metadata.exe window.

The Issue Separation menu contains three functions that assist with issue folder creation and image file sorting. The New Meta hotkey creates a new metadata.txt file for an issue folder open in Windows Explorer. The Help menu for every TDNP application contains a link to the software documentation on the Digital Projects Unit wiki.

## Chapter 3: Super\_Metadata

### Super\_Metadata Templates

Super\_Metadata templates add general descriptive information to the metadata records for groups of similar newspaper issues during the upload process for the Portal to Texas History. In order to create , the DNU collects information for the following metadata fields:

- **Main Title:** The main title reflects the actual title printed on the front page. The main title also includes the place of publication in parentheses. For example: The Cherokeean Herald (Rusk, Tex.).
- **Serial Title:** The serial title reflects the title listed in the U.S. Newspaper Directory.
- **Creator:** If available, include the newspaper's editor as the creator. List the editor in the following format: Last Name, First Name. Use authorized terms from the Library of Congress name authorities whenever possible: <http://authorities.loc.gov/>
- **Publisher:** If available, record the newspaper's publisher. Include the place of publication and any additional information about the publisher, such as an office address, in the publisher information field.
- **Content Description:** Include a descriptive statement about the newspaper's content.
- **Physical Description:** Include the original newspaper's physical dimensions and layout elements according to Library of Congress standards.
- **Subject:** List the appropriate standard subject terms and any additional keywords terms in the subject field. The example below illustrates the standard subjects for a newspaper published in Sweetwater, Texas. The DNU lists the following Library of Congress Subject Headings (LCSH) and University of North Texas Libraries Browse Structure (UNTL-BS) subjects for each newspaper:
  - i. (UNTL-BS) Business, Economics and Finance – Communications - Newspapers
  - ii. (UNTL-BS) Business, Economics and Finance – Journalism
  - iii. (UNTL-BS) Business, Economics and Finance – Advertising
  - iv. (UNTL-BS) Places – United States – Texas – Nolan County
  - v. (LCSH) Nolan County (Tex.) -- Newspapers.
  - vi. (LCSH) Sweetwater (Tex.) -- Newspapers.
  - vii. (LCSH) Sweetwater (Tex.) -- Newspapers.

- **Coverage:** Include the geographic and time period coverage information for the newspaper. Formulate geographic coverage terms according to the Getty Thesaurus of Geographic Names listing: <http://getty.edu/research/tools/vocabularies/tgn/> The Digital Projects controlled vocabularies outline the various time period categories to choose from: <http://digital2.library.unt.edu/vocabularies/coverage-eras/>
- **Collection:** Select the appropriate collections from the Digital Projects controlled vocabulary: <http://digital2.library.unt.edu/vocabularies/collections/> Each record usually includes the Texas Digital Newspaper Program collection and a collection for the newspaper's title if there are a substantial number of issues available. If the newspaper's digitization was funded by a grant, include the grant's collection name.
- **Institution:** Select the institution that contributed the newspapers from the Digital Projects controlled vocabulary: <http://digital2.library.unt.edu/vocabularies/institutions/>
- **Language:** List all languages found in the newspaper.
- **Identifier:** List the newspaper's LCCN and OCLC numbers available on the newspaper's US Newspaper Directory record.
- **Note:** Include a **Display Note** for additional descriptive information about the newspaper. Include your name and the date you created the `super_metadata.xml` template in a **Non-Display Note**.

## New Record Creator

The DNU uses a Web-based tool called the New Record Creator (Figure 3.1) to create and export the Super\_Metadata descriptive metadata to an XML file. Follow the steps below to create `super_metadata.xml` files for a newspaper title:

1. Create folders for each year represented in the title, and move all issue folders to the appropriate year folder. For example, move all issues published in 1957 to a new folder named 1957. Add the Reel Number to the name of each report (e.g. UNT\_0620-report-2012-12-12.txt). Move all reports to a new folder called reports.
2. Gather the descriptive metadata, as described in the Super\_Metadata Templates section above, for the first issue in the first year folder.
3. Open the New Record Creator in a Web browser, and select Create a New Record: <http://edit.texashistory.unt.edu/nrc/>
4. Fill in the descriptive metadata for the first issue in the appropriate fields. Refer to the UNTL metadata guidelines for further information and instructions: <http://library.unt.edu/digital-projects-unit/input-guidelines-descriptive-metadata>

5. Select “Yes” in the Primary Source field and “Visible” from the radio buttons in the Tools sidebar.
6. After entering the descriptive metadata, click the Export Record button and save the file as `super_metadata.xml` in the year folder that contains the first issue.

The screenshot shows the 'New Record Creator' interface for a record titled 'NT Daily'. The interface is divided into several sections for entering metadata:

- TITLE:** Contains three title fields:
  - Main Title:** North Texas Daily (Denton, Tex.)
  - Serial Title:** North Texas Daily
  - Alternate Title:** NT Daily
- CREATOR:**
  - name:** Gorman, Sean
  - type:** Personal
  - role:** Editor
  - info:** Editor-in-chief
- PUBLISHER:**
  - name:** University of North Texas
  - location:** Denton, Texas
  - info:** (Empty field)

On the right side, there is a **Tools** sidebar with the following options:

- Records based off of this template should have an initial state of:
  - Hidden
  - Visible
- Buttons: Export Record, Save Record as Template, Clear/Reload Template
- Field counts:
 

Title (3)	Citation (0)
Creator (1)	Relation (0)
Contributor (0)	Collection (2)
Publisher (1)	Institution (1)
Date (0)	Rights (0)
Language (1)	Resource Type (1)
Description (2)	Format (1)
Subject (16)	Identifier (2)
Primary Source (1)	Degree (0)
Coverage (2)	Note (1)
Source (0)	
- Collapsible All Fields

Figure 3.1: The New Record Creator editing interface.

After exporting the initial `super_metadata.xml` file, use the New Record Creator's import feature to create `super_metadata.xml` files for the rest of the newspaper title's year folders. Follow the steps below to create additional `super_metadata.xml` files:

1. Select "Import A Record" from the Templates menu in the upper left corner.
2. Click the Browse button to locate the first `super_metadata.xml` file.
3. Click the Submit button to display the metadata (Figure 3.2).
4. Compare the first and last issues in each year folder for changes in the metadata. Common changes include a new editor or publisher and subtle changes to the information recorded in the Publisher Info and Display Notes.
5. **If the metadata changes**, click the Edit/Create Template button, input the new metadata in the New Record Creator, and complete steps 6 and 7 below.
6. Create a new year folder. Rename the year folders to reflect the chronological sequence. For example, name the first folder `1957-1` and the second folder `1957-2`.
7. Export the edited `super_metadata.xml` file from step 5 to the new folder.
8. **If the metadata remains the same from one year to the next**, simply copy and paste the previous year's `super_metadata.xml` file into the next year's folder. Each year folder must contain a `super_metadata.xml` file.

Figure 3.2: The New Record Creator's Import page with a record displayed.

## Chapter 4: Metadata Post-Processing

### Portal to Texas History Dashboard

The Portal to Texas History Dashboard provides a Web-based interface that the DNU uses to edit published metadata records. We occasionally discover errors in published newspaper records. The Dashboard provides a convenient interface to input the correct metadata. The Dashboard editing interface is similar to the New Record Creator editing interface (see Figure 3.1 in the previous chapter). To change a published record's metadata, perform the following steps:

1. Find the record in the Portal to Texas History public interface.
2. Type `edit.` at the beginning of the record's URL and press the Enter key.
3. Make the necessary corrections to the metadata in the editing interface.
4. Click the Publish button.

### Selenium IDE Automation

The DNU uses Web automation software called Selenium IDE to make batch changes to the metadata for newspapers published on the Portal to Texas History. Metadata errors often occur across multiple issues in a newspaper title. If the required corrections are standard, the DNU uses Selenium IDE to quickly make those corrections for large batches of metadata records.

Follow these general instructions to locate and edit batches of records:

1. Use the Dashboard's search feature to isolate the set of records that require corrections.
2. Record a test case in Selenium IDE, and create a test suite to enable batch automation.
3. Load multiple records as tabs in the Firefox browser window.
4. Play the test suite to automatically edit, publish, and close the records.

More detailed instructions about how to use Selenium IDE to make batch metadata corrections are available on the Digital Projects Unit wiki at the following URL:

[http://digitalprojects.library.unt.edu/projects/index.php/PracticalMetadata\\_Selenium](http://digitalprojects.library.unt.edu/projects/index.php/PracticalMetadata_Selenium)





## Conclusion

This document describes the process that the University of North Texas Libraries Digital Newspaper Unit uses to create high quality metadata for the digital newspapers published on the Portal to Texas History. The four steps of the process include: Metadata Production, Metadata Quality Control, Super\_Metadata, and Metadata Post-Processing. This manual presents the workflows and standards as of the time of writing. For the most current information and additional tools and techniques, please visit the Newspapers section of the Digital Projects Unit wiki page at the following URL:  
[http://digitalprojects.library.unt.edu/projects/index.php/Project\\_Listings#Newspapers](http://digitalprojects.library.unt.edu/projects/index.php/Project_Listings#Newspapers)

## References

- Digital Projects Unit Wiki*. (2012). University of North Texas Libraries. Retrieved from:  
<http://digitalprojects.library.unt.edu/projects/index.php>
- Getty Thesaurus of Geographic Names*. (2012). The Getty Research Institute. Retrieved from:  
<http://getty.edu/research/tools/vocabularies/tgn/>
- Input Guidelines for Descriptive Metadata*. (2012). University of North Texas Libraries. Retrieved from:  
<http://library.unt.edu/digital-projects-unit/input-guidelines-descriptive-metadata>
- Library of Congress Authorities*. (2012). The Library of Congress. Retrieved from: <http://authorities.loc.gov/>
- Portal to Texas History*. (2012). University of North Texas Libraries. Retrieved from:  
<http://texashistory.unt.edu>
- RLG Guidelines for Microfilming to Support Digitization*. (2003). Research Libraries Group. Retrieved from:  
[http://oclc.org/resources/research/publications/library/Pres\\_Micro\\_Supplement.pdf](http://oclc.org/resources/research/publications/library/Pres_Micro_Supplement.pdf)
- Texas Digital Newspaper Program*. (2008). University of North Texas Libraries. Retrieved from:  
<http://tdnp.unt.edu>
- University of North Texas Digital Projects: Controlled Vocabularies*. (2012). University of North Texas Libraries. Retrieved from: <http://digital2.library.unt.edu/vocabularies/>
- US Newspaper Directory*. (2012). The Library of Congress. Retrieved from:  
<http://chroniclingamerica.loc.gov/search/titles/>



## Appendix: Newspaper Metadata Glossary

---

Term	Definition
AutoHotkey	Open source software for Windows that allows users to create keyboard shortcuts and applications to automate tasks. Information about how to create AutoHotkey scripts is available on the Digital Projects Unit wiki at the following URL: <a href="http://digitalprojects.library.unt.edu/projects/index.php/HowToOther_AHK">http://digitalprojects.library.unt.edu/projects/index.php/HowToOther_AHK</a> The DNU's applications for the Texas Digital Newspaper Program are located in the following folder: Q:\Software\AutoHotKey\TDNP_Apps
Collation	Process of gathering data about microfilmed newspaper issues. The DNU records the number of pages for each issue as well as information about problems such as missing or duplicate page images. See also Evaluation.
Content Description	Metadata element that describes what a newspaper is about. This is an example of the standard formulation used by the DNU: Daily newspaper from Sweetwater, Texas that includes local, state and national news along with advertising.
Controlled Vocabulary	Set of terms with standard definitions used to facilitate retrieval of information objects. The complete UNT Digital Projects Controlled Vocabulary is available at the following URL: <a href="http://digital2.library.unt.edu/vocabularies/">http://digital2.library.unt.edu/vocabularies/</a>
Creator	Metadata element that indicates who is responsible for the newspaper's content. If available, list the newspaper's editor(s) in the following format: LastName, FirstName.
Collection	Metadata element that groups similar newspapers together to facilitate retrieval. The complete list of UNT Digital Projects Collections is available at the following URL: <a href="http://digital2.library.unt.edu/vocabularies/collections/">http://digital2.library.unt.edu/vocabularies/collections/</a>
Coverage	Metadata element that describes the newspaper's geographic and time period scope. Formulate new geographic coverage terms according to the Getty Thesaurus of Geographic Names listing: <a href="http://www.getty.edu/research/tools/vocabularies/tgn/">http://www.getty.edu/research/tools/vocabularies/tgn/</a> The complete list of UNT Digital Projects Coverage time periods is available at the following URL: <a href="http://digital2.library.unt.edu/vocabularies/coverage-eras/">http://digital2.library.unt.edu/vocabularies/coverage-eras/</a>
Dashboard	Online editing tool developed by the UNT Digital Projects Unit which the DNU uses to edit published metadata records: <a href="http://edit.texashistory.unt.edu/">http://edit.texashistory.unt.edu/</a>
Description	Metadata element that provides information about the newspaper's content and physical characteristics. See also Content Description and Physical Description.
Evaluation	Process of determining the physical and content attributes of a microfilm roll. Careful microfilm evaluation helps the DNU to create high quality images and metadata for its digital newspapers.
Identifier	Metadata element that provides the unique numbers associated with a particular newspaper in the Library of Congress and OCLC catalogs. See also LCCN and OCLC.
Institution	Metadata element that lists the organization that provided the microfilmed newspaper for digitization. The complete list of UNT Digital Projects Institutions is available at this URL: <a href="http://digital2.library.unt.edu/vocabularies/institutions/">http://digital2.library.unt.edu/vocabularies/institutions/</a>
Issue Folder	Folder that contains the images, issue metadata, and OCR files for one newspaper issue. The issue folder name provides issue date metadata.

Issue Metadata	Text file (metadata.txt) that contains the Volume and Issue numbers and additional information about the issue in a Note.
LCCN	Library of Congress Control Number. Unique identifier associated with a specific newspaper title. The LCCN is available in the newspaper's US Newspaper Directory Record: <a href="http://chroniclingamerica.loc.gov/search/titles/">http://chroniclingamerica.loc.gov/search/titles/</a>
Language	Metadata element that shows the language(s) contained in the newspaper. A newspaper may contain more than one language.
Main Title	Metadata element that lists the title printed on the front page of the newspaper. The Main Title may be different for newspapers with the same Serial Title. For example, the word "The" may be added to or removed from the beginning of a newspaper's title.
Metadata	Information about a digital newspaper. The DNU produces metadata in three locations: the issue folder name, the metadata.txt file, and the super_metadata.xml file.
Microfilm	Preservation format for historical newspapers. The DNU digitizes most of its newspapers from second generation 35 mm. microfilm negatives.
Microfilm Reader	Device used to view microfilm images. The image is projected onto a screen with an incandescent bulb.
MSYS	UNIX-like shell environment for Windows. The DNU uses the MSYS shell and a Python script to produce data about the issues on a microfilm roll. See also Report.
New Record Creator	Online editing tool developed by the UNT Digital Projects Unit which the DNU uses to produce super_metadata.xml templates for groups of similar newspapers: <a href="http://edit.texashistory.unt.edu/nrc/">http://edit.texashistory.unt.edu/nrc/</a> See also Super_Metadata.
Newspaper Notes	Controlled Vocabulary for the Notes section of the metadata.txt file. The complete Newspaper Notes standard is available on the Digital Projects Unit wiki at the following URL: <a href="http://digitalprojects.library.unt.edu/projects/index.php/Newspaper_Notes">http://digitalprojects.library.unt.edu/projects/index.php/Newspaper_Notes</a>
OCLC Number	Unique identifier associated with a specific newspaper title. The OCLC number is available in the newspaper's US Newspaper Directory Record: <a href="http://chroniclingamerica.loc.gov/search/titles/">http://chroniclingamerica.loc.gov/search/titles/</a>
OCR	Optical Character Recognition. Software that analyzes a newspaper image and produces a text file for full-text searching.
Physical Description	Metadata element that provides information about a newspaper's physical characteristics. This is an example of the standard formulation used by the DNU: pages : ill. ; page 21 x 17 in. Digitized from 35 mm. microfilm.
Portal to Texas History	Web site hosted by the University of North Texas Libraries where the DNU publishes its digital newspapers: <a href="http://texashistory.unt.edu/">http://texashistory.unt.edu/</a>
Primary Source	Metadata element that indicates whether or not an information object contains original material. All digital newspapers are primary sources.
Publisher	Metadata element that indicates who produced the newspaper. If available, list the name of the newspaper's publisher (individual or organization) as it appears on the masthead or an inside page.
Q: Drive	Network hard drive that houses the DNU's digital newspaper files.
QC	Quality Control. Process that ensures high quality images and metadata for the DNU's digital newspapers.

---

Reel	Plastic device, similar to the rim of a tire, that holds a microfilmed newspaper. Also referred to as a Roll.
Reel Folder	Folder that contains the image files produced from a reel of microfilm. Also referred to as a Roll Folder.
Report	Text file produced after images have been separated into issue folders that the DNU uses for issue metadata quality control. Instructions for creating reports are available on the Digital Projects Unit wiki at the following URL: <a href="http://digitalprojects.library.unt.edu/projects/index.php/TDNP_IssueCheck.py">http://digitalprojects.library.unt.edu/projects/index.php/TDNP_IssueCheck.py</a>
Selenium IDE	Web browser automation software for Firefox. The DNU uses Selenium IDE to make batch changes to published metadata records in the Dashboard. Instructions for using Selenium IDE are available on the Digital Projects Unit wiki at the following URL: <a href="http://digitalprojects.library.unt.edu/projects/index.php/PracticalMetadata_Selenium">http://digitalprojects.library.unt.edu/projects/index.php/PracticalMetadata_Selenium</a>
Serial Title	Metadata element that indicates the newspaper's title as listed in the US Newspaper Directory: <a href="http://chroniclingamerica.loc.gov/search/titles/">http://chroniclingamerica.loc.gov/search/titles/</a>
Subject	Metadata element that includes controlled vocabulary terms and keywords that describe what a digital newspaper is about. See page 15 of the Newspaper Metadata Manual for more information.
Super_Metadata	XML file that provides descriptive metadata for groups of similar digital newspapers. The information contained in the super_metadata.xml file is added to the metadata record for each issue when digital newspapers are uploaded to the Portal to Texas History.
UNTL-BS	University of North Texas Libraries – Browse Structure. Controlled vocabulary for Subject terms developed by the Digital Projects Unit. The complete UNTL-BS list is available at the following URL: <a href="http://digital2.library.unt.edu/subjects/?formId=main&amp;fieldId=subject">http://digital2.library.unt.edu/subjects/?formId=main&amp;fieldId=subject</a>
US Newspaper Directory	Web site maintained by the Library of Congress that contains catalog records for United States newspapers from 1690 to the Present: <a href="http://chroniclingamerica.loc.gov/search/titles/">http://chroniclingamerica.loc.gov/search/titles/</a>

---