

Digital newspaper preservation through collaboration

Digital
newspaper
preservation

Ana Krahmer

University of North Texas, Denton, Texas, USA

73

Received 3 September 2015
Accepted 28 September 2015

Abstract

Purpose – The Texas Digital Newspaper Program (TDNP) supports newspaper preservation and access for any title in Texas, from any date, any location and representing any community. As an active member of the Texas Press Association, TDNP also supports large-scale preservation of born-digital newspaper PDF issues for member publishers. This paper aims to explore how the early days of TDNP built a strong foundation of collaboration and support for large-scale preservation projects, including support for preserving a state press association PDF newspaper collection.

Design/methodology/approach – This paper is a case study of a collaborative endeavor to create a large-scale, statewide digital newspaper preservation hub in Texas. This paper details how individual partnerships led to new and larger partnerships. Figures and tables represent numbers of partner institutions served, numbers of newspapers preserved and screenshots of how these items appear within collections on the digital repository environment of The Portal to Texas History. This paper concludes with recommendations for groups interested in developing their own collaborative projects.

Findings – As a case study, the data explored include numbers of partnering institutions, materials contributed by partnering institutions and how these numbers help to forward the TDNP agenda.

Practical Implications – The final recommendations are lessons learned through collaboration, and the implications are real-world advice from the partners developed through the TDNP.

Originality/value – Hosting over 3.25 million pages of newspapers, the TDNP has become an enormous hub of newspaper preservation in Texas, and it is unique in the numbers of partners it supports and the numbers of pages it is able to host for free access via The Portal to Texas History. This paper is intended to help other groups across the world build their own collaborative preservation efforts, and it offers pragmatic advice derived from hands-on experience.

Keywords Collaboration, Partnership, Digital libraries, Newspapers, Digital preservation, Historical research

Paper type Conceptual paper

The Texas Digital Newspaper Program (TDNP) supports newspaper preservation and access for any title in Texas, from any date, any location and representing any community. As an active member of the Texas Press Association (TPA), TDNP also supports large-scale preservation of born-digital newspaper PDF issues for member publishers. This article will explore how the early days of TDNP built a strong foundation of collaboration and support for large-scale preservation projects, including support for preserving a state press association PDF newspaper collection.

History and background: responding to community needs

In the early 2000s, Cathy Nelson Hartman, now Associate Dean of UNT Libraries, envisioned partnering with communities across Texas who hold primary source history materials to develop what has now become The Portal to Texas History: a unified, statewide digital repository for primary source artifacts. University of North Texas



Libraries established the Portal in 2002 to help improve access to the rich archival resources held at cultural heritage institutions across Texas. During outreach visits with small and large partner organizations, a common theme emerged of communities that especially wanted to digitize and preserve their local newspapers. Dean Hartman discussed her experiences from when she spoke with different Texas groups on these outreach visits:

As we visited many cities and towns across Texas talking about our vision for the Portal, talking with them about their collections and the value of the collections for researchers, the first and most important content on their list for digitization was their local newspapers. The newspapers contained the history of their communities – local events, announcement of births/deaths/weddings, news of the city governments, local politics, etc., and sometimes included a local view on national and international news. (C. Hartman, personal communication, May 1, 2013).

Newspapers were a digitization priority for communities, but newspaper digitization required a strong technological infrastructure combined with staff expertise to maintain high digitization standards.

National digital newspaper program

To build staff knowledge, UNT Libraries, in collaboration with the Dolph Briscoe Center from American History at the University of Texas, was awarded the National Digital Newspaper Program (NDNP) grant in 2007:

NDNP, a partnership between the National Endowment for the Humanities (NEH) and the Library of Congress (LC), is a long-term effort to develop an Internet-based, searchable database of USA newspapers with descriptive information and select digitization of historic pages. Supported by NEH, this rich digital resource will be developed and permanently maintained at the Library of Congress (NDNP, 2013).

UNT Libraries leveraged this funding opportunity as a springboard toward developing a more robust newspaper preservation program for the Portal and for Texas:

We knew that it was important that we develop a plan for digitizing and preserving newspapers and including them in the Portal. The NEH funding helped us move this initiative forward (C. Hartman, personal communication, May 1, 2013).

During the 2007 award cycle of the NDNP program, UNT staff seized the opportunity to work with national standards, technological and infrastructural requirements, and staff workflow needs.

The scope of NDNP is very specific in terms of title, year and location selections. As a result, titles to be digitized through NDNP were to be “titles of note” selected by an advisory committee of state historians, archivists and librarians from available microfilm holdings that had not previously been digitized. The first NDNP award cycle originally covered the years of 1880-1910: “In subsequent award years, the time period was gradually extended decade by decade, to cover the historic period of 1836-1922” (2013). The funding from *Chronicling America* supported digitization of newspaper content from early, large city Texas dailies which the board chose for their high historic value. During the first funding cycle, UNT Libraries would add 100,000 pages of newspapers to *Chronicling America*, as well as to *The Portal to Texas History*.

The Texas digital newspaper program

Upon public announcement of UNT Libraries as a partner institution in NDNP, Portal staff were immediately inundated with requests from partners from across the state to digitize various titles that fell completely outside the scope of NDNP, including titles from small towns, post-1923 issues of titles and, in some instances, titles that were in languages other than English. To meet with these requests, the TDNP began. Portal staff spread the message that any and all Texas newspaper content was fair game for TDNP, whether it came from a town of 600 people or whether it represented Fort Worth during its cattle boom. The goal was that TDNP would serve as the hub of preservation and access needs to digital newspaper content in Texas.

The TDNP was conceived to promote standards-based digitization and preservation of newspapers throughout Texas. Through continual outreach visits across Texas combined with advanced technological infrastructure and multiple funding sources, TDNP has now become a state- and nation-wide leader in newspaper preservation and digitization. Portal staff also developed a website (<http://tdnp.unt.edu>) and workflow for collaborating with institutions interested in digitizing their communities' newspapers, with the goal of broadening the scope of the NDNP within a repository dedicated to Texas history materials.

Technology and workflows

The Portal to Texas History now manages three content streams for newspaper materials: microfilm digitization, paper digitization and born-digital ingests. This work is carried out by a team consisting of four full-time staff members and six student assistants. Once content has been digitized and organized by the newspaper team through metadata and page sequencing, it undergoes Optical Character Recognition (OCR) using an Abbyy Recognition Server cluster comprising 10 nodes and 48 OCR core. OCR enables full-text searchability, with the search hits appearing as highlighted in yellow (Figure 1, for an example of a search result for "Sam Houston"). To date, newspapers of nine different languages have been OCR'ed and included on The Portal to

SAM HOUSTON SOCIETY.
The **Sam Houston** Literary Society feels justly proud of its progress this year. Its membership, about forty in number, exceeds all previous records, and the earnest work of the members is very gratifying. And although the intermediate debate was rendered impossible by some of the participants being called away too late for their places to be filled, yet great hopes are entertained for the inter society debate, and it is not feared but that the reputation of the society will not be sustained. It is greatly deplored that there will be no opportunity for measuring strength with outside contestants this year as both societies are in good condition and would doubtless reflect credit upon the school. One thing however is not as it should be, with the coming of Spring the attendance is becoming irregular. This will never do. If we expect to compete successfully with other schools we must keep up our society work, and if we hope to enter college we cannot afford to neglect the training that we are able to receive here.

Spalding's Sporting Goods
The Best
Only to be found at Richardson's

BOYS, Drink Water from the
GEORGETOWN Mineral Wells
It will make you well. 50 per gallon at wells, 100 per gallon delivered at your home.
H. BURGARDT (Owner) Proprietor.

Let The
Troy Steam Laundry
Do Your Laundering.
Good Work and Low Prices.
ROHWELL, KEY & LANGSTON

Figure 1.
Example search for "Sam Houston", highlighted on page 2 of the Friday, March 13, 1908, issue of The Megaphone

Texas History. In addition to OCR, every newspaper issue loaded into TDNP has an accompanying metadata record to ensure access points to the issues.

UNT Libraries purchased a high-end microfilm scanner in 2011 with the plan to internalize much of the newspaper digitization process to lower the cost for partners and control more of the workflow. This purchase allowed the Portal staff to retool workflows to optimize the throughput of newspaper pages being digitized. During this time, the UNT Libraries operated a microfilm-only digitization shop. This meant that if a partner approached the Portal to digitize content that was only in paper format, the workflow involved sending the newspaper content to the Oklahoma Historical Society (OHS), where it would be microfilmed. After masters and second-generation negative copies were made at OHS, the Portal staff would digitize this content in its existing microfilm workflow.

By 2013, UNT Libraries invested in a new, large-format planetary scanner to allow direct digitization from paper newspapers in a cost- and time-effective manner. With this additional scanner, there is no need to scan from microfilm, though some partnering libraries request microfilm duplicates of their newspapers from OHS for archival preservation purposes. The file quality levels to which physical newspaper pages are scanned results in much larger file sizes than those produced from microfilm (Table I) because of full color content and larger dots or pixels per inch displayed in digital format. Images scanned from physical pages of newspaper display at a much higher quality than microfilm because they never underwent reduction ratio conversion (see Figures 2 and 3 for a comparison between newspapers scanned from microfilm and from physical pages).

Building partnerships

One of the founding goals of The Portal to Texas History is to unite content providers across Texas with UNT Libraries' capacity for digitization services, digital preservation infrastructure and open data access. Groups who partner with UNT Libraries to digitize their newspapers frequently cite the lack of fees for continuing access or continued sustainability to their communities as a deciding factor in their decision to digitize with UNT Libraries. One library director whose group had received a grant to digitize and preserve their newspaper in TDNP stated:

The newspaper digitization project places our library in position to reach out toward the future. By taking a piece of the past and bringing it with us, we are sure to grow and learn, appreciate and respect what was, what is and what will be (L. Whitaker, personal communication, January 24, 2014).

A large portion of partners who contribute materials to TDNP are small community public libraries who cannot necessarily afford continuing sustainability fees.

Image types	Bit depth	Color space	Resolution (dpi)	Scale	File format
Microfilm	8-bit	Grayscale	400 ^a	100% (1:1)	Tiff 6.0 (uncompressed)
Physical pages	24-bit	RGB	400	100% (1:1)	Tiff 6.0 (uncompressed)

Table I.
TDNP file standards
for scanning
newspapers

Note: ^aFor microfilm, scans are acquired at 8-bit grayscale with a resolution of 400 dpi, if possible, otherwise between 300 and 400 dpi (relative to the size of the original newspaper)



Figure 2.
Example wagon
advertisement from
Brenham Evening
Press, August 20,
1910, scanned from
microfilm



Figure 3.
Example wagon
advertisement from
the *Honey Grove
Signal*, September 2,
1910, scanned from
physical page

All newspaper issues on the Portal are assigned two organizational units: a “Partner” and a “Collection”. We define “Partners” as those groups who have organized funding for the digitization project. We define “Collection” as a contextual whole that represents newspaper dataset. “Collection” could represent newspapers from multiple towns within one county, or it could represent a series of newspapers and photographs that all come from the same city. Many partners also include non-newspaper content on The Portal to Texas History, and the “Partner” field lets partners combine their data under the umbrella of their institutional biography while still separately representing the different data types that exist in other collections under their institution.

For example, the University of Texas at El Paso has placed 3,389 newspaper issues on the Portal, along with 45 other objects that reside in different Collections, including a City Directories collection and a collection containing primary sources from the Mexican Revolution (Figure 4). All newspapers are also always included in the TDNP collection.

Data about the numbers of new partners added each year (Table II) directly feed into content acquisition goals for the long term and short term. Because most newspaper digitization projects for the Portal are funded according to a grant cycle schedule, as previously mentioned, we use this information to seek out prospective new partners who might be interested in applying for grants to fund digitization of their newspaper content.



Figure 4. Items and collections from the University of Texas at El Paso, as they appear on The Portal to Texas History

Year	No. of partners added	No. of new partners added	Cumulative no. of partners
2006	2	2	2
2007	0	0	2
2008	2	2	4
2009	6	5	9
2010	10	6	15
2011	16	12	27
2012	20	10	37
2013	24	10	47
2014	52	32	79
2015	46	14	90

Table II. New partners added to TDNP per year

Often partners from different types of institutions have distinctive patterns in newspaper content holdings (Table III). For example, genealogical/historical societies and museums will hold a variety of regional newspaper titles that range in date from very early to mid-twentieth century, but they rarely have complete newspaper runs. Academic libraries often have a combination of regional newspapers along with a complete run of their student newspaper, and the student newspapers demonstrate a severe drop in page count during the summer months. Public libraries frequently either own an entire run of their community’s newspaper or have direct access to their respective newspaper publishing offices.

Data about numbers of issues contributed by each partner (Table IV) give the TDNP team a good way to encourage partners to digitize more newspaper content and to apply for newspaper digitization grants from local foundations and state agencies. These data are easily viewable from the TDNP collection statistics page (<http://texashistory.unt>).

[edu/explore/collections/TDNP/stats/](#)). Current partners report that they use their collection and institution statistics pages to determine how much more of their newspaper content they might digitize, while prospective partners utilize these pages to see which groups in their region have already contributed newspaper content to TDNP and as a deciding factor in whether they also want to contribute their newspapers. For example, multiple institutions in the panhandle region of Texas contributed one after another, after communicating with each other about the process and after using the Newspaper Program statistics pages.

Some partners, like the Dolph Briscoe Center for American History, have entered into multiple large-scale projects with TDNP to add a wide variety of different newspapers. After successfully partnering with the Dolph Briscoe Center for American History on the NDNP grant, they continued to partner with UNT Libraries through a project called *The Early Texas Newspapers Project* (<http://texashistory.unt.edu/explore/collections/ETXNP/>), funded by a Texas State Library and Archives Commission grant program, TexTreasures. The titles digitized for this project represent most of the earliest Texas newspapers available on The Portal to Texas History.

Starting in 2008, multiple collaborations have formed between UNT Libraries and other Texas higher education institutions ([University of North Texas Libraries, 2008](#)). Early partnerships of note include work with the Abilene Library Consortium and its member institutions who were interested in digitizing the school newspapers from the higher education institutions they represent. At that time, a total of 10,384 issues were digitized from 15 titles of four educational institutions. The Abilene Library Consortium newspapers, now standing at over 44,000 issues, serve as a cornerstone of The Portal to Texas History and provided Portal staff with the opportunity to improve upon digitization workflows in building TDNP. In addition to the Abilene Library Consortium newspaper content, UNT Libraries partnered with Rice University to add its student newspaper, the *Thresher*, comprising 36,248 news pages. TDNP staff have also worked with Texas Wesleyan University and the University of Dallas to digitally preserve and make accessible their born-digital PDF content on the Portal. Acquired from physical pages is the St. Mary’s University student newspaper, *The Rattler*, contributed by the Louis J. Blume Library. Student newspapers in Texas are frequently not microfilmed, and the ability of TDNP to acquire content from physical page and PDF issues helps fill a preservation gap in Texas.

Many participating newspaper publishers and local public libraries are in the process of adding and preserving entire runs of their titles on The Portal to Texas History. In 2010, the Singletary Memorial Library in Rusk, Texas, partnered with their publisher to begin the process of digitizing the complete *Rusk Cherokeean Herald*, starting with its first issue in 1919 and moving into the present day. In 2011, the Bastrop Public Library began digitization of *The Bastrop Advertiser*, starting with issues from 1854 and

Partner type	No. of partners by type	No. of issues
Public libraries	52	108,402
Academic libraries and archives	25	174,683
Genealogical/historical societies	8	14,518
Museums	5	14,422
Private collectors	1	2

Table III.
Number of TDNP
partners and
newspaper issues
contributed by
institution type

Name	Issues contributed (items)	Partner type
Abilene Christian University Library	3,461	Academic
Abilene Library Consortium	44,599	Academic
Abilene Public Library	9,042	Public library
Archer City Public Library	150	Public library
Atlanta Public Library	3,652	Public library
Bartlett Activities Center and the Historical Society of Bartlett	3,355	Genealogical/Historical societies
Bastrop Public Library	5,428	Public library
Baylor County Free Library	1,004	Public library
Bellaire Friends Library & Historical Society	1,827	Genealogical/Historical societies
Boyce Ditto Public Library	252	Public library
Breckenridge Public Library	11,914	Public library
Bryan Wildenthal Memorial Library (Archives of the Big Bend)	115	Public library
Carrollton Public Library	417	Public library
Carson County Library	315	Public library
Casey Memorial Library	1,893	Public library
Collin County Genealogical Society	2,259	Genealogical/Historical societies
Crosby County Public Library	3,243	Public library
Deaf Smith County Library	1,651	Public library
Dennis M. O'Connor Public Library	623	Public library
Denton Public Library	90	Public library
Dr Hector P. Garcia Memorial Library	243	Public library
Dublin Public Library	3,842	Public library
Eastland Centennial Memorial Library	6,668	Public library
Ellis Memorial Library	1,471	Public library
Ferris Public Library	155	Public library
Gaines County Library	2,280	Public library
Gibbs Memorial Library	1,888	Public library
Gladys Johnson Ritchie Library	207	Public library
Grayson County Frontier Village	7,009	Museum
Hardin-Simmons University Library	2,557	Academic
Harrie P. Woodson Memorial Library	2,381	Public library
Hemphill County Library	3,051	Public library
Heritage House Museum	1 item	Museum
Higgins Public Library	1,131	Public library
Hondo Public Library	1,956	Public library
Honey Grove Preservation League	1,365	Genealogical/Historical societies
Howard Payne University Library	6,474	Academic
Hutchinson County Library, Borger Branch	1,914	Public library
Jacksonville Public Library	47	Public library
Kerens Public Library	314	Public library
Lamar State College-Orange	8,438	Academic
Lamar University	2,005	Academic
Lancaster Genealogical Society	754	Genealogical/Historical societies
Livingston Municipal Library	7	Public library
Lubbock High School	90	Academic
Lucy Hill Patterson Memorial Library	976	Public library

Table IV.
TDNP partner
institutions and
number of
newspapers
contributed by each
(continued on
following pages)

(continued)

Name	Issues contributed (items)	Partner type
McMurry University Library	1,858	Academic
Meridian Public Library	512	Public library
Mesquite Public Library	1,722	Public library
Mineola Memorial Library	1,413	Public library
Montgomery County Memorial Library	62	Public library
Newton County Historical Commission	89	Genealogical/Historical societies
Palacios Library	1,245	Public Library
Panola College	445	Academic
Private Collection of Mary Newton Maxwell	2	Private collection
Reagan County Library	1,435	Public library
Rice University Woodson Research Center	3,055	Academic
Richard S. and Leah Morris Memorial Library	1,702	Public library
Rio Grande City Public Library	1,888	Public library
Sachse Public Library	408	Public library
Schulenburg Public Library	1,780	Public library
Shamrock Public Library	3,550	Public library
Singletery Memorial Library	4,569	Public library
Slovanska Podporujici Jednota Statu Texas	4,260	Genealogical/Historical societies
Smith County Historical Society	609	Genealogical/Historical societies
Smith Public Library	1,882	Public library
South Texas College of Law	277	Academic
Southwestern University	2,402	Academic
St. Mary's University Louis J. Blume Library	184	Academic
St. Mary's University School of Law	25	Academic
Stella Hill Memorial Library	1,984	Public library
Stonewall County Library	2,982	Public library
Sweetwater/Nolan County City-County Library	5,030	Public library
Swisher County Library	3,976	Public library
Tarleton State University	2,555	Academic
Texas State University	525	Academic
Texas Wesleyan University	1,186	Academic
The 12th Armored Division Memorial Museum	778	Museum
The Colony Public Library	66	Public library
The Dolph Briscoe Center for American History	13,491	Academic
The Grace Museum	119	Museum
The Old Jail Art Center	6,515	Museum
UNT Libraries	72,169	Academic
UNT Libraries Special Collections	1,471	Academic
University of Dallas	245	Academic
University of Texas at El Paso	3,389	Academic
University of Texas at San Antonio	3,707	Academic
University of Texas at San Antonio Libraries Special Collections	381	Academic
West Public Library	2,461	Public library
Western Texas College Library	139	Academic
Zula B. Wylie Memorial Library	1,415	Public library

Note: This list is also visible from the TDNP collection statistics page, in the tab labeled "More data"

Source: University of North Texas Libraries (2015)

Table IV.

digitized up to 2013. In 2012, the Crosby County Library and its publisher collaborated to digitize the entire run of the *Crosby County Review*. Also in 2012, the Sweetwater City-County Library and their publisher began digitizing the *Nolan County News* as well as *The Sweetwater Reporter*, which upon completion will consist of almost 100,000 pages and 120 years. The partnerships between public libraries, publishers and the TDNP illustrate how collaboration can lead to digital preservation.

The Texas press association

As TDNP and its partnerships have expanded, the need to shape the identity of the program through public communication has increased tremendously. This need is a direct result of more newspaper preservation work, which has led to increased workload, increased staff and directly correlates to the need for greater financial support. Shaping the identity of TDNP has involved a series of conversations with external stakeholders and internal administrators. This process has included such work as developing better online information about partnering with TDNP (<http://tdnp.unt.edu>), as well as creating and deploying traveling newspaper banners to reside at area public libraries and schools. Also, this process has involved raising awareness of TDNP through participation in library conferences and TPA annual meetings.

Since 2012, UNT Libraries have held an active membership in the TPA, which allows the Libraries to communicate the importance of newspaper preservation and access. Publishers who meet with TDNP representatives at annual TPA meetings have approached UNT Libraries to learn more about working with them. To tailor the message to publishers, TDNP strives to emphasize the value of digitally preserving newspapers. Many publishers are concerned about incurring a financial loss should they make their newspapers freely available; therefore, UNT Libraries gives publishers the option to preserve their newspaper content but to embargo it for a designated number of years so that it is preserved but not publicly accessible. Terrie Gonzalez, publisher of *The Rusk Cherokeean-Herald*, says:

Newspapers have long been recognized as a mirror, reflecting the lives of those being served. The printed word maintains today a strong position for the reflection of people and their lives, perhaps more so today than ever before. The historical value is capsuled in these words: "We must look to the past to appreciate the present and plan for the future" (T. Gonzalez, personal communication, August 26, 2012).

Ms Gonzalez preserves the PDF editions of her newspaper and embargoes the most recent three years on The Portal to Texas History.

TDNP staff often speak with publishers who believe that the best business opportunities lay in monetizing their historic content, and they do so by working with private vendors to establish paywalls to their newspaper content. This approach can create problems, particularly in terms of preservation: often many corporate vendors working with these newspapers do not necessarily have the technological or financial infrastructure to maintain large newspaper archives according to preservation imaging standards. Additionally, because corporations so frequently shut down, there is some risk in believing that perpetuity with a vendor means the same thing as perpetuity with a state-run academic library. In discussing these preservation issues with publishers, Portal staff always stress the importance of scanning or preserving PDF born-digital content, according to Library of Congress standards.

It is no accident that Cathy Hartman and UNT Libraries’ digital newspaper team heavily contributed to the “NDSA Case Study for Newspaper E-Prints”, which recommends educating stakeholders like publishers, educators, researchers, public libraries and historical societies through conferences, annual meetings, workshops and presentations; this education would help spread the message of newspaper PDF print masters as being just as important as microfilming physical newspapers. Connecting communities with their publishers to digitize newspapers has been a highly successful means toward digital newspaper preservation, and UNT Libraries have completed many successful newspaper projects by practicing what they preach.

Born-digital newspaper preservation

Born digital-newspapers serve as a significant content ingest stream for TDNP. This began in 2010 when Terrie Gonzalez contacted Portal staff to archive born-digital, print master versions of her newspaper, they were able to work with her to understand the specifics of born digital newspapers and develop a workflow that met the needs of both the publisher and the Portal. After successfully completing this first born-digital newspaper project, the Portal staff was poised to add a class of newspaper content into its workflows for newspaper digitization. Both the born-digital newspapers and the newspapers scanned from the large-format planetary scanner have the added benefit of providing full-color content to Portal users.

By the early summer of 2014, the TPA met with TDNP staff to discuss the possibility of digitally preserving the TPA PDF Archive, which is a deposit archive for publishers to preserve their born-digital, PDF print master content. The TPA Archive will thus be stored in two locations: one, with Newz Group in Missouri and the other with UNT Libraries. By August 2014, this resulted in nearly 1.5 million pages of Texas newspapers in PDF format, dating from 2010 to present. TDNP staff have reached out to publishers in Texas to communicate with them about embargo time periods. The default time period for embargo is 96 years, after which the objects will be in the public domain. Often, when publishers learn that the default option is 96 years, they tend to request shorter embargo periods, of between six months and five years, with six months being the most common embargo period. Publishers explain to TDNP staff that they choose six month periods because that is the amount of newspaper content they host behind a paywall for their paying subscribers.

PDF issues that are embargoed will appear within the “Collection” statistics page (see Figure 5), listed as “Invisible” and “Hidden”, but at present, users cannot view what

Statistics for Texas Digital Newspaper Program

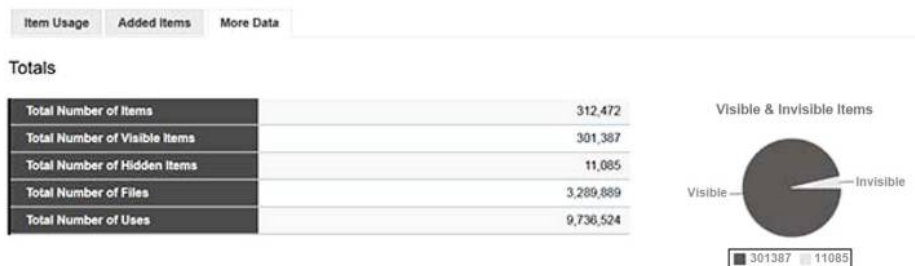


Figure 5. TDNP statistics display on The Portal to Texas History, showing visible and hidden items

issue dates are hidden or what title pages may have looked like. In the future, Portal staff would like to make bibliographic metadata about the preserved issues visible to the public when the digital content is not available. This is planned for a future revision of The Portal to Texas History interface.

Research projects of note

Multiple research projects utilizing TDNP newspaper content demonstrate how a state hub of newspaper digitization can support the larger research conversation.

One specific example of a newspaper title that led to a highly significant project is the *Representative*, from Galveston. According to Diana Kleiner of the Texas State Historical Association, “The first black newspaper printed in Texas, the *Representative* was published by editor and proprietor Richard Nelson at Galveston from 1,870 to June 1,872” (Kleiner, 2010, n.p.). TDNP staff found this newspaper title amongst Harvard’s microfilm holdings, and since its addition to the Portal, historians like Dr Elizabeth Turner have studied it on The Portal to Texas History in preparation of new research articles. Dr Turner has based one chapter of a new book on this newspaper, titled, “Three Cheers for Freedom and Equal Rights to All”, which explores Richard Nelson’s editorials about the Juneteenth festivities of the early 1870s.

Another project that heavily utilizes digitized Texas newspapers comes from Dr Andrew Torget of the University of North Texas and Jon Christensen of Stanford University’s Bill Lane Center for the American West. This project, “Mapping Texts”, was designed as:

[...] a collaborative project between the University of North Texas and Stanford University whose goal has been to develop a series of experimental new models for combining the possibilities of text-mining and geospatial analysis in order to enable researchers to develop better quantitative and qualitative methods for finding and analyzing meaningful language patterns embedded within massive collections of historical newspapers (Torget, 2014, p. 3).

This project utilized the raw OCR data from approximately 230,000 newspapers to draw out patterns and meaning in Texas history. This project provided a starting point for other historical research projects that explore TDNP data to explore larger historical patterns.

The Texas Runaway Slave Project out of the East Texas Research Center at Stephen F. Austin University similarly utilizes OCR data from TDNP newspapers to depict the lives of runaway slaves in the South during the mid-nineteenth-century. Special Collections Librarian Kyle Ainsworth, out of the East Texas Research Center at Stephen F. Austin University, is the primary researcher behind this project:

The Texas Runaway Slave Project (TRSP) began in December 2012 at Stephen F. Austin State University. Since then, more than 10,000 Texas newspaper issues, published prior to 1,865, have been indexed, from which the information and materials for the project will be gathered. The majority of these issues are found in two databases: The Texas Digital Newspaper Project [*sic*] (<https://tdnp.unt.edu/>) and the Portal to Texas History (<http://texashistory.unt.edu/>) (Ainsworth, 2014).

These newspapers contain biographical information about Texas slaves whose identities have not been filled in over time. As a result, the runaway slave project helps teach new researchers about who these people were and how they lived.

A follow-up project to the Texas Runaway Slave Project is the Texas Runaway Ads project, a joint-endeavor between UNT History Department Andrew Torget's graduate students and Rice University History professor Caleb McDaniel's undergraduate students. Torget, McDaniel and their students have collaborated on extensive research on newspapers available in TDNP to build the Texas Runaway Ads (@TxRunawayAds, <https://twitter.com/TxRunawayAds>) Twitter feed displays actual runaway slave advertisements that appeared in newspapers published between 1836 and 1860, with the idea of representing the plight of Texas runaway slaves who otherwise might not have had representation in other formats, such as through biographical essays or printed oral history transcripts. The geographic regions these students focused on were centers of agricultural commerce in Texas during the mid-nineteenth century, and their newspapers are some of the earliest available in TDNP. Set within the university environment, this project has taught students a great deal, especially about the trade skills and educational background that Texas slaves had to perform their duties (Torget, 2014, personal interview).

Research like this is possible through the array of digital newspapers that have been preserved and made openly accessible for research of any kind, and this preservation process is possible through the contributions and missions of multiple funding sources.

Recommendations and conclusion

This paper is intended to represent a model of how successful collaborative efforts have contributed to large-scale preservation of a state's newspaper content. Individual strands of thread seem very small, but combining strands forms something much larger and more meaningful. The TDNP serves as an example of how combining the threads of resources, skillsets and content can lead to something much bigger and more all-encompassing than what would have occurred if UNT Libraries had determined that the Portal would be a stand-alone digitization effort.

Concluding recommendations for building strong collaborations to support long-term endeavors, based on the example that TDNP provides, are:

- *Match what your institution has to what other institutions need:* In the case of TDNP, UNT Libraries do not have 3 million pages of Texas newspapers in microform or physical pages. Working with content holders who did not necessarily have the resources to digitize their newspapers, UNT Libraries were able to trade the resources they had, technological infrastructure and skillsets, for the content they needed.
- *Start small:* When beginning a collaborative effort, TDNP staff recommend a pilot project to potential partner institutions. This way, both groups can get a sense for what each will contribute to the overall project and each group can see how their contribution represents a portion of the sum of all parts.
- *Start local:* This advice applies to groups who need funding to begin collaborative projects. TDNP stands as proof that local granting foundations are often very supportive of local projects, and research at places like Foundation Center and its directory (<http://foundationcenter.org/>) really pays off by providing information about what a given foundation supports. If there is interest in collaboration but the project needs funding, the collaborating institutions should begin by applying

for support with a smaller, local foundation rather than focusing only on large-scale grantors, such as NEH, Knight or IMLS grants. Smaller pilot projects supported through local foundations also provide good leverage for representing success metrics, and collaborative institutions applying to much larger granting agencies can use these metrics to apply for larger grants as a next step in the collaborative process.

- *Expect the unexpected:* Something TDNP staff learned is that groups who were not involved in early collaboration became very interested in participating when they realized how the Texas newspapers in TDNP were used, and usable, in research projects. Although it is the job of UNT Libraries and content contributors to make newspapers available and to preserve them for the long term, everyone involved is always amazed by the types of information people find, and by the variety of research questions people have answered, from the digital newspaper collection. TDNP has surprised all of its contributors in how much it has grown to be larger than the sum of its parts.
- *Spread the word:* Travel funding at academic institutions is becoming increasingly scarce, but a large part of the success of TDNP can be attributed to the willingness of UNT Libraries to send Portal and TDNP staff to events and conferences where they may represent their work and encourage participation. When one potential partner obtains grant funding to digitize a newspaper after having attended a TDNP presentation, the travel fee pays itself back.

The Portal to Texas History and the TDNP stand as a testament to collaboration and long-term planning around technological infrastructure, content and services. These projects not only encompass local preservation and access to the rich resources in the State of Texas but also help serve as models for other states and organizations who are interested in developing a similar operation for their users. Newspaper content, both current and historical, plays an important role in helping us understand our pasts, and it is therefore an important component in cultural heritage digitization projects. Successfully organizing and leveraging community support for the digitization and preservation of this content that is scattered, the state provides an important foundation for future projects, research and learning.

References

- Ainsworth, K. (2014), "Methodology", The Texas Runaway Slave Project, available at: <http://digital.sfasu.edu/cdm/methodology/collection/RSP>
- Kleiner, D.J. (2010), "Galveston County", *Handbook of Texas Online*, available at: www.tshaonline.org/handbook/online/articles/hcg02
- Torget, A. (2014), "Texas Runaway Ads", available at: <https://twitter.com/TxRunawayAds>
- University of North Texas Libraries (2008), "TexTreasures: Early Texas newspapers: 1829 – 1861", available at: <http://texashistory.unt.edu/explore/collections/ETXNP/>
- University of North Texas Libraries (2015), "The Texas digital newspaper program", The Portal to Texas History, available at: <http://texashistory.unt.edu/explore/collections/TDNP/>

Further reading

Cayley, S. (2012), "Digitization in teaching and learning", *Victorian Periodicals Review*, Vol. 45 No. 2, pp. 210-214.

Library of Congress (2012), "National digital newspaper program", *Chronicling America*, available at: <http://chroniclingamerica.loc.gov>

National Digital Information Infrastructure and Preservation Program (2013), "About – Digital preservation", available at: www.digitalpreservation.gov/about/

National Digital Stewardship Alliance (2013), "A case study: Newspaper e-prints", available at: www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_CaseStudyNewspaperEPrints.pdf

Corresponding author

Ana Kraemer can be contacted at: ana.kraemer@unt.edu