

Laying the Groundwork for Newspaper Preservation through Collaboration and Communication: The Texas Digital Newspaper Program

Ana Krahmer
Mark Phillips
University of North Texas Libraries

Abstract

University of North Texas Libraries established the Texas Digital Newspaper Program (TDNP) to digitize any Texas newspaper title, of any date, and to digitally preserve and make them available via The Portal to Texas History. Through site visits to multiple Texas libraries and personal interviews with librarians, genealogists, educators, students, and historians, UNT Libraries prioritized newspaper digitization within the content scope for The Portal to Texas History and determined processes for acquiring and ingesting multiple formats of newspapers, including from physical papers, microfilm, and born-digital PDF print masters.

UNT Libraries gained the technological education to lay the underpinnings on which to host and preserve Texas newspapers through their work with the National Endowment for the Humanities' National Digital Newspaper Program (NDNP). While the scope of the NDNP program is to digitize pre-1923, public-domain Texas newspapers, TDNP seeks to make any Texas date and title freely accessible. As a result, generating support through communication with newspaper publishers, public libraries, and academic libraries is crucial. In addition, communication and collaboration are critical to locating Texas newspapers because Texas faces a unique problem in that many of its newspapers, whether in microfilm or physical format, are scattered amongst multiple institutions and private vendors.

This presentation will elaborate on the financial, communicational, and technological processes involved in building the Texas Digital Newspaper Program. UNT Libraries digitally preserves and makes freely available, via The Portal to Texas History, over 1 million pages of Texas newspapers, spanning from 1829 to the present. The Texas Digital Newspaper Program is a case study in digital preservation and open access to digitized newspapers and is utilized by multiple communities of users, including genealogists, academic and lay historians, and K-12 and university researchers.

Introduction

The Texas Digital Newspaper Program (TDNP), hosted on The Portal to Texas History out of University of North Texas (UNT) Libraries, is the hub of newspaper digitization and preservation in Texas. At present, TDNP hosts over 1 million pages of Texas newspapers, digitally preserved and made openly available to the world, with issue-level metadata description and full-text search capability. The process of building TDNP has involved extensive community outreach, internal and external financial support, and technological evolution to create a viable preservation and usable access infrastructure from which to host TDNP.

Texas faces a somewhat unique problem in the United States when it comes to digitally preserving and making its newspapers freely accessible: multiple institutions and corporate vendors each own different Texas newspaper titles on master microfilm and in physical paper format. While some states, like Oklahoma, began building a physical newspaper archive under one institution very early in the twentieth-century, Texas newspapers are scattered across three large collecting institutions and three corporate vendors; in addition, many local public libraries, archives, and historical societies hold the only copies of their respective newspaper titles.

History & Background

The Portal to Texas History: Responding to Community Needs

In the early 2000s, Cathy Hartman, now Associate Dean of UNT Libraries, envisioned working with the 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 to Texas History in 2002 to help improve access to the rich archival resources held at cultural heritage institutions around the State of Texas. UNT Libraries laid the groundwork for the Portal through the support of a competitive grant from the Telecommunications Infrastructure Fund, a pool of money set aside at that time to assist in the implementation of technology infrastructure in libraries, schools and museums around the state. During the initial award period the first technical infrastructure for the Portal was designed and built by an open-source software company in Denmark, IndexData. In addition to infrastructure, technology and standards adoption, the project conducted outreach with partners throughout the State of Texas to create a base set of content providers and interested parties in this newly developing portal.

During outreach visits with small and large partner organizations, a common theme emerged, which was the desire amongst communities to digitize and preserve their local newspapers. Dean Hartman discussed her experiences from when she spoke with different groups on these 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). Lynda James-Gilboe well summarizes this universal desire for access to historic newspapers: "Growing interest in primary source material, an increasing impatience (or occasionally, disdain) for manual searching, space savings, the need to preserve and archive in multiple formats, and the quickly changing expectations of a new generation of researchers accustomed to immediate access to electronic resources are driving an expanding need for newspaper digital archives" (James-Gilboe 2005). However, the idea of newspaper digitization was daunting to Portal staff: so much rich textual content would require tremendous server space, strict graphics standards, OCR licensing, and good metadata access points. For these reasons, newspaper digitization was an area The Portal to Texas History staff initially did not feel confident about pursuing.

In 2005 The Portal to Texas History launched, and the first collections were opened to the public, containing photographs, personal papers, books--all primary source materials that represented life in Texas history. During the next years everything about the Portal grew: the number of partners, the number of collections, the number of items in the system, and most importantly the number of users of the Portal grew. As the user-base began to recognize the value of the aggregation of content offered by The Portal to Texas History, the staff of the Portal at UNT began to hear more questions about when they would add historic newspapers to the system. "Historical newspapers present several unique digitization challenges: large image size, complex formats that change from page to page, and stories that are continued on another page or pages. These factors make scanning more difficult. Headlines seldom appear in the tightly packed text of turn-of-the century newspapers, making it tricky to identify individual articles. Limited space between columns presents an issue as does content that was filmed while bound" (2005). Notwithstanding, the Portal staff gained the confidence to try digitizing newspapers, and this time the Portal partnered with the Ferris Public Library, which had received external grant funding to digitize and preserve their town's paper, *The Ferris Wheel*.

Through this project, UNT Libraries digitized 219 issues of this newspaper from the original paper and hosted the content on The Portal to Texas History. This small pilot allowed Portal staff to experiment with new ways of delivering the large-format newspaper content via the Web, and helped us understand the complexities, like those James-Gilboe highlights, that would be involved in moving forward with newspaper content. Next, we determined to seek additional content and training, through the National Endowment for the Humanities' National Digital Newspaper Program.

National Digital Newspaper Program

In 2007 UNT Libraries, in partnership with the Dolph Briscoe Center from American History at the University of Texas Austin, applied for the National Digital Newspaper Program (NDNP) award: "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 U.S. 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" (2013). UNT was selected as the lead institution in Texas to coordinate the newspaper digitization activities for this program. The Portal staff leveraged this funding opportunity as a springboard toward developing a more robust newspaper program for the Portal and the State of Texas. "We knew that it was important that we develop a plan for digitizing 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 learn more about and engage standards, technological and infrastructural requirements, and staff workflow needs.

The scope of NDNP is very specific in terms of title, year, and location selections. The Library of Congress' Chronicling America website explains its scope selection and changes over each following cycle: "Each NDNP participant receives an award to select and digitize approximately 100,000 newspaper pages representing that state's regional history, geographic coverage, and events of the particular time period being covered. In order to plan for phased development, the annual award program began with targeting digitized material for the decade 1900-1910" (2013). 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 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.

Texas Digital Newspaper Program

Upon public announcement of University of North Texas Libraries as a partner institution in the NDNP program, the 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. Seth Cayley, in his article, "Digitization in Teaching and Learning," succinctly explains the same paradox we saw, of newspaper content we digitized through NDNP versus the content other groups also wanted to include on The Portal to Texas History: ". . .there is a danger of assuming that what is not digitized is unimportant" (2012). For the Portal, we decided to spread the message that we were interested in digitizing all Texas newspaper content, whether it came from a town of 600 people, or whether it represented Fort Worth, Texas, during its cattle boom. To further this message, UNT Libraries organized the Texas Digital Newspaper Program, to serve as the hub of preservation and access needs to digital newspaper content in the State of Texas.

After the National Digital Newspaper Program and the Ferries Wheel projects, the Texas Digital Newspaper Program (TDNP) was born: to promote standards-based digitization of newspapers throughout the State of Texas. The Texas groups with whom Dean Hartman spoke almost unanimously reported that newspaper digitization was a high priority to them. 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.

To prepare the message for TDNP's purpose, the Portal staff developed a website and workflow for collaborating with institutions interested in digitizing their communities' newspapers. Portal staff wanted to broaden the scope of the NDNP through TDNP because there remained a considerable number of newspapers which we did not digitize through the NEH grant, for example: newspapers from very small communities; newspapers in languages other than English; newspapers that contain only a small number of known issues which did not represent full runs of the paper; newspapers that only

published for a short period of time; newspapers published after 1923; and student newspapers from colleges and universities. The number of newspapers out of scope of the NDNP program was indeed larger in scope than the number of eligible titles that were in scope. TDNP serves as a vehicle for Texas to digitize its vast newspaper collections in a cost-effective, highly-usable way.

External Funding, Internal Success, and Extensive Partnerships

TDNP and the External Community

Through on-site presentations with multiple community groups from across Texas, listserv announcements, and constant and active communication, the Portal team got the message across to Texas libraries about newspaper preservation and digitization. Multiple state groups began to fund UNT to add more newspaper titles to The Portal to Texas History. In 2008, UNT Libraries partnered again with the Dolph Briscoe Center for American History to apply for a TexTreasures grant from the Texas State Library and Archives Commission. This grant was awarded to UNT over three successive cycles, for the Early Texas Newspapers Project, which digitized over 58,000 pages of Texas newspapers, spanning 1829-1952, because of the significance of the collection:

Early Texas newspapers are vital resources for understanding the history of Texas during its colonial period, republic era, and early statehood. Because few of these newspapers have been microfilmed or digitized, scholars must travel to collections that house them and handle the fragile original copies. In order to enhance the availability of early newspapers to historians, researchers, and the public, the University of North Texas (UNT) Libraries and the Center for American History (CAH) at the University of Texas intend to partner to microfilm, digitize, and provide free public access via the Internet to the earliest Texas newspapers held by CAH, many of which have never been microfilmed before. (2008)

Another Texas-based foundation, the Tocker Foundation, has established very close ties to UNT Libraries and has awarded a total of \$463,339 to public libraries from across Texas who serve communities with populations below 12,000. The Tocker Foundation is heavily involved in improving access to free and open information through rural public library services. As a result, the Foundation's aspirations closely connect to those of UNT Libraries for The Portal to Texas History. Through the Tocker Foundation's tremendous support, over 315,000 pages of newspapers are openly accessible on the Portal, and another 100,000 are undergoing the digitization process as this article is being written. These 415,000 pages of newspapers represent eighteen communities from across Texas whose public library directors have applied to receive digitization grant funding to make their newspapers available on The Portal to Texas History. These pages comprise almost 38% of the total newspaper pages on the Portal. The active participation of small-population public libraries in the Tocker Foundation's Newspaper Digitization Program demonstrates the value that communities place on their own history and on the need to create access to their newspapers, and it illustrates the strong reputation of TDNP and the Portal. As these libraries work through the process of applying for Tocker Foundation grants, they also establish or strengthen ties to their own newspaper publishers.

TDNP: Internal to Campus

In 2012, University of North Texas released an updated strategic plan, called the *Four Bold Goals for the Future*, intended as a roadmap for how the university would grow and evolve from 2012-2017. The Portal to Texas History particularly fits well with two of these four goals because its foundation is based upon collaborative outreach for digital preservation of community identity according to strict, high standards.

One of UNT's goals is to become a national leader in student support, to offer students the best possible access to educational resources in the country. Because the Portal serves as a free and open gateway to a wide variety of primary source materials, including newspapers, UNT faculty, graduate, and undergraduate students have heavily utilized the materials, both in the classroom and in their personal research. For example, Dr. Andrew Torget, a faculty member in the UNT Department of History, assigns a specific project in his History of Texas course which requires students to choose and perform research on TDNP newspapers (A. Torget, personal communication, May 14, 2013). Because of this assignment's heavy focus on 19th-century Texas history, students extensively utilize the Early Texas Newspaper project, which represent the earliest newspapers on the Portal. According to Dr. Torget, students complete the assignment with surprise at how much more they learn about the world from the first-hand newspaper articles rather than from history textbooks.

Another of the university's goals that closely parallels the goals of the Portal is to collaborate as a regional leader of mutually "beneficial partnerships and resources" (2013). The heavily outreach-oriented work of the Portal and TDNP also well represents the university goal to create a base of support for the region. The highly collaborative processes involved in building a gateway to primary source materials of the extent that the Portal has become ensure digital preservation. The National Digital Information Infrastructure and Preservation Program, begun in 2000, defines itself as, "based on an understanding that digital stewardship on a national scale depends on public and private communities working together" (NDIIPP 2013). It is no coincidence that Cathy Hartman, an original visionary of the Portal, was strongly involved in the National Digital Information Infrastructure and Preservation Program in the early 2000s and has continued her work with it through the National Digital Stewardship Alliance. The emphasis of UNT's goal on collaboration comes to life in the work of the UNT Libraries on building Portal to Texas History collections like the Texas Digital Newspaper Program.

Significant Partnerships with TDNP

From 2008-2013, multiple partnerships formed between UNT Libraries and higher education institutions from across Texas. 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. In total 10,384 issues were digitized from 15 titles of four educational institutions. These Abilene Library Consortium newspapers serve as a cornerstone of The Portal to Texas History and provided us with the opportunity to improve upon digitization workflows in building TDNP. In addition to the Abilene Library Consortium newspaper content, we have partnered with Rice University to add its student newspaper, the *Thresher*, comprising 25,592 pages. We 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. The digital newspaper team is in the process of including

the entire UNT newspaper run, *The Campus Chat* and *NT Daily*, in the Portal, which began publication on November 1, 1916, and continues to the present day.

From across Texas, multiple 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 its most recent PDF print masters. In 2011, the Bastrop Public Library began digitization of *The Bastrop Advertiser*, starting with issues from 1854, and digitized up to the present-day PDF print masters. 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 Texas Digital Newspaper Program illustrate how collaboration can lead to digital preservation.

Communicating the Texas Digital Newspaper Program

Identity

TDNP has grown through many successful collaborative partnerships with the goal of preserving newspapers. As TDNP has grown, the need to shape the identity of the program through public communication has increased tremendously. This need is a direct result of more newspaper work, which has led to increased workload, increased staff, and directly correlates to the need for greater financial support. Toward the end of constructing the TDNP message, UNT Libraries hired a technical communicator, Ana Krahmer, in 2011 to supervise the digital newspaper team, which offered the opportunity for rhetorical planning in preparing the TDNP message. Shaping the identity of the Texas Digital Newspaper Program has involved a series of conversations and with external partners with internal administrator. 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 for a period of 6-8 weeks at area public libraries and schools. Also, this process has involved raising awareness of TDNP through participation in conferences and with the Texas Press Association.

As of 2012, UNT Libraries holds an active membership in the Texas Press Association, which allows the Libraries to communicate the importance of newspaper preservation and access. As a result of the newspaper team's participation at the annual Texas Press Association conference, publishers have approached UNT Libraries to learn more about working with them. To tailor our message to publishers, we strive to emphasize the value of digitally preserving newspapers, downplaying the idea of free and open access. Many publishers are concerned about incurring a financial loss should they make their newspapers freely available; therefore, we give publishers the option to preserve their born digital PDF print masters 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 print masters of her newspaper and embargoes the most recent three years on The

Portal to Texas History; because the Portal is both a preservation and access medium, it is possible to flip the switch on each following year to open the content to the world while still preserving the dark content.

In Texas, we will frequently 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 can create other obvious 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 come and go in this day and age, 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, we stress the importance of scanning or preserving PDF content, according to LC standards. Cathy Hartman and UNT Libraries' digital newspaper team heavily contributed to the recent "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 has completed many successful newspaper projects by practicing what they preach.

Technology: Evolution and Redesign

Testing and Upgrades

In 2008, UNT Libraries received an Institute of Museum and Library Services (IMLS) grant to perform usability research on a very specific target audience, genealogists, toward the goal of improving overall user experience of The Portal to Texas History. This grant, titled *IOGENE: Interface Optimization for Genealogists*, allowed us iteratively to test and design for a non-university population, genealogists, who would heavily traffic the Portal, if the Portal could respond to their needs through testing and iterative re-design: "The IOGENE project will identify a model for digital library interface development using a user-centered design approach. Genealogists will be involved in the design process beginning with an assessment of their information use environment and continuing through usability testing of the redesigned Portal interface" (IOGENE 2013). As a result of this series of focus groups, surveys, and user testing, the IOGENE project improved very specific elements of newspaper access for users, even while it provided the financial assistance necessary to redesign the Portal's technical infrastructure.

In 2009 the UNT Libraries upgraded the underlying infrastructure used by The Portal to Texas History and other systems. To respond to user testing and feedback, we changed key elements to newspaper usability, including building a new interface for viewing newspapers which allowed for zooming into each page, highlighting search terms on the page, and finally printing articles and sections of interest from the paper:

Our system currently uses zoomify (Flash based) to view different maps and photographs in our system that we have at high detail. We also have a function for books that allows us to highlight words that people search for within each book we have in our collection. We are now in the process of digitizing newspapers for our collection, and wanted to combine those two features

for users viewing our newspapers. The problem is that zoomify doesn't do this, and we don't currently program in Flash. So we found a Javascript alternative to zoomify called Brainmaps, which isn't quite as nifty as zoomify, but was easier for me to modify, and already had a system similar to highlighting called "labeling". It also uses the same jpeg tile structure that zoomify requires, so it will be easier to migrate all our previously tiled digital objects into our new system. (Fredericks 2008)

The work we did in 2008 to improve usability of the zoom and highlight functionalities of our newspapers exemplifies the goal of the project to re-use existing technology from other communities whenever possible as a way of extending the resources for our project. In this example the technology we finally adopted is a combination of the OpenLayers map viewing system that allows the use of the Zoomify tile structure for displaying different tile levels to the end user. An open source Python tool is used for the conversion of TIFF and JPEG2000 images into the tiled JPEG structure used by Zoomify, a single WARC file is created which holds the 2000+ images associated with each page of newspaper. The infrastructure upgrade also allowed the Texas Digital Newspaper Program and the Portal to ingest larger amounts of content into the system and to make this content available to the end users much more quickly than in the years past.

Present Technology and Workflows

The Portal to Texas History now manages four content streams for newspaper materials: microfilm digitization, paper digitization, born-digital ingests, and finally content digitized using the Library of Congress NDNF standards and specifications. This work is carried out by a team consisting of four full-time staff members and four student assistants. Work is overseen by the assistant dean of Digital Libraries at UNT, Mark Phillips. Once content has been digitized and organized by the newspaper team, it undergoes Optical Character Recognition (OCR) using an Abbyy Recognition Server cluster comprised of 10 nodes and 48 OCR cores. UNT Libraries maintains a license of 3 million pages per year for this server and aims to utilize the license fully by the end of each year. To date newspapers of seven different languages have been OCR'ed and included on The Portal to Texas History.

In 2011 UNT Libraries purchased a high-end microfilm scanner with the plan to internalize much of the newspaper digitization process in order 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 programmed and microfilmed. After masters and second-generation negative copies were made at OHS, the Portal staff would digitize this content in its existing microfilm workflow.

The microfilm workflow remained in place until the Spring of 2013, when UNT Libraries made the decision to invest 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 are now two workflows in place for digitization of analog newspaper content, one from paper and another from microfilm.

Born digital-newspapers also factor heavily in the Texas Digital Newspaper Program. When in 2010 Terrie Gonzalez, publisher of the Rusk *Cherokee Herald*, contacted the Portal staff to archive born-digital, print master versions of her newspaper, we 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.

Identification, Collection, Digitization, Preservation and Access

Our workflows are such that all staff on the digital newspapers team and the head of Digital Libraries collaborate to balance out the workload across the entire process of adding newspapers to the Portal. Because of this collaboration, all work does not take place at one person's level; instead, everyone does a little bit of something, as represented below.

Identification is often done by the head of the digital newspaper team, and this process begins with:

- Speaking with communities interested in digitization.
- Verifying content availability, either in microfilm, physical pages, or born-digital PDFs.
- Requesting publisher permission.
- Locating funding sources.

Collection refers to the process of getting the content one permissions and funding are in place, and this process is usually performed by the head of the digital newspaper team:

- Microfilm: locate master microfilm to create and purchase 2nd-generation, duplicate negatives (2N) copies of the newspaper.
- Physical newspapers: arrange for transport of physical pages to the UNT Libraries for scanning.
- Born-digital content: provide publisher or partnering library with an external hard drive onto which they can copy print master newspapers.

Preservation, performed across all staff on the digital newspaper team, refers to the process of digital preservation by which all newspapers included in the Texas Digital Newspaper Program are maintained, including:

- Ingesting scanned or born-digital images and accompanying issue information into the preservation server.
- Describing with README information that identifies Who, What, When, Where, How the materials originated and came to be loaded into The Portal to Texas History.
- Preserving in perpetuity, to guarantee both access and newspaper fidelity.

Access, achieved through work by all digital newspaper team members, refers to the means by which users can read digitized newspapers on the Portal:

- Descriptive metadata is created and available at the issue-level.
- Newspaper images are viewable as downloadable JPGs in extra small, small, medium, large, and extra-large sizes.
- Optical Character Recognition provides full-text search.
- Zooming allows users to read newspapers close-up for easier visibility.

Future Directions & Conclusion

The Portal and the Texas Digital Newspaper Program have multiple goals for the future. These include expansion to include more Texas titles from larger communities and cities, which will involve negotiating with publishers about copyright. The Portal staff would like to conduct further usability research to study how users navigate between issues of newspapers, and to see how they read individual newspapers--whether they read cover-to-cover or spot-read for specific names. We hope to work with more publishers to preserve PDF print masters because these represent a growing and significant collection in need of digital preservation.

In the last six years, the Portal has added over 1.2 million pages of Texas newspapers to its repository. These pages constitute 133,000+ issues of newspaper from 550 titles. There are over 135 communities from 65 of the 254 Texas counties represented in this collection. Since 2009 there have been over 3.5 million item uses of this newspaper content, and as more content is added to the system the usage increases. While the first million newspaper pages online was a noteworthy milestone for TDNP and the Portal, it is just the first step in the identification, collection, preservation and access to these vibrant primary source items from the State of Texas.

Works Cited

- Cayley, S. (2012). Digitization in teaching and learning. *Victorian Periodicals Review*, 45(2), 210-214.
- Federicks, B. (2008, March 31). Django's cache middleware. [Web log post]. Retrieved May 15, 2013, from <http://iogene.library.unt.edu/?p=45>.
- James-Gilboe, L. (2005). The Challenge of digitisation: Libraries are finding that newspaper projects are not for the faint of heart. *Serials Librarian* 49(1/2). 155-163.
- Library of Congress. (2012). National digital newspaper program. *Chronicling America*. Retrieved May 15, 2013, from <http://chroniclingamerica.loc.gov>.
- National Digital Information Infrastructure and Preservation Program. (2013). About - Digital preservation. Retrieved May 13, 2013, from <http://www.digitalpreservation.gov/about/>.
- National Digital Stewardship Alliance. (2013). A Case study: Newspaper e-prints. Retrieved May 28, 2013, from:
http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_CaseStudy_NewspaperEPrints.pdf.
- University of North Texas. (2012). Four bold goals for the future of one great university. *University of North Texas Strategic Plan 2012-2017*.
- University of North Texas Libraries. (2013). About IOGENE: Interface optimization for genealogists. Retrieved from http://iogene.library.unt.edu/?page_id=29.
- University of North Texas Libraries. (2008). *TexTreasures: Early Texas newspapers: 1829 – 1861*.