CREATING CITIZEN HISTORIANS

JOHN HERBERT AND KAREN ESTLUND

Historical newspaper research has become increasingly straightforward with digitization and has allowed historical research to be more accessible for the general public. The Utah Digital Newspapers program is a seminal project that has led the way for libraries and archives to digitize their historical newspapers. With over five million newspaper articles online, first-hand reports of historical events are at the fingertips of anybody with an Internet connection.

Newspapers are the first draft of history, primary sources of historical information used by scholarly and lay researchers alike. Researchers typically access historical newspapers in central locations like libraries, and they must be diligent enough to browse reels of microfilm or read aged originals. With time and effort, these methods can be effective, but they have never been efficient. They simply are too slow and tedious to create a broad audience for newspaper research.

The advent of digitized newspapers, available and *searchable* on the Internet, is fundamentally changing this paradigm. With a growing number of historic newspapers coming online, anyone with an Internet-connected PC can conduct thorough newspaper research. Consequently today, in the twenty-first century, newspaper digitization is exploding.

The J. Willard Marriott Library at the University of Utah (U of U), partnering with Brigham Young University (BYU) and Utah State University (USU), runs a pioneering state-wide program that is changing the face of newspaper research. Since its inception in 2002, the Utah Digital Newspapers (UDN) program quickly became a model for other academic libraries across the country and federal agencies in Washington, D. C. We have been, and remain, on the leading edge for newspaper digitization, especially within the public sector.

JOHN HERBERT is the Program Director for the Utah Digital Newspapers and Head of Digital Technologies at the J. Willard Marriott Library, University of Utah. At this writing, he is also serving as interim director of the University of Utah Press.

KAREN ESTLUND, Digital Collections Coordinator at the University of Oregon Libraries, managed the Utah Digital Newspapers Program during 2006–2007 in her former position at the University of Utah's J. Willard Marriott Library.
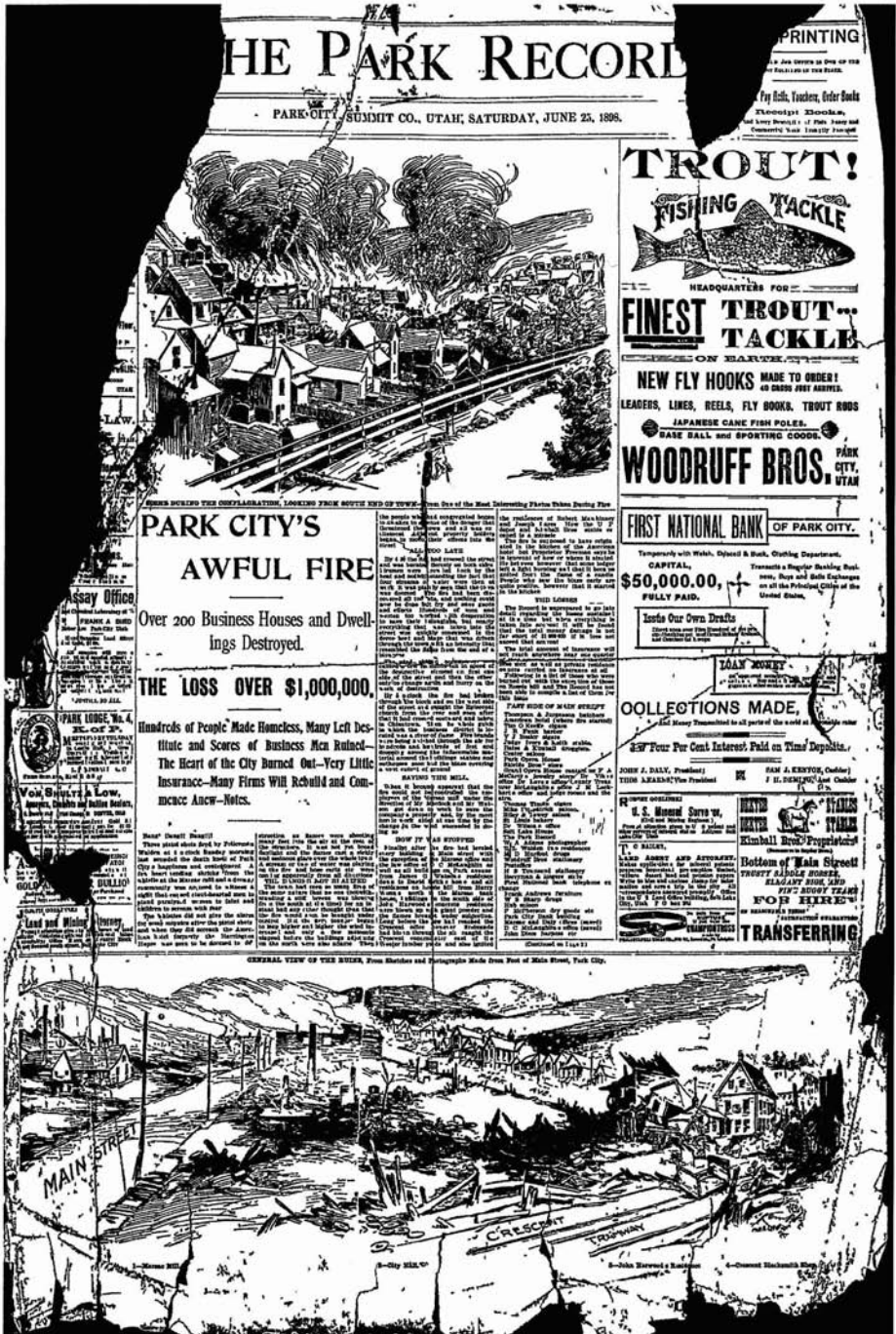
Figure 1. *Park (Park City, UT) Record*, describing "Park City's Awful Fire" on 25 June 1898. This article can be found in the University of Utah's J. Willard Marriott Library's Utah Digital Newspapers Collection at http://udn.lib.utah.edu/u?/park,39400.

Our program loads digital copies of historic Utah newspapers onto our Web site, making them available to the general public over the Internet. Anyone with an interest in history can search them by keyword or browse them by title and date, all from the comfort of their PC. UDN is easily accessible, easy to use, and is beginning to make reading newspaper microfilm obsolete. You can find our Web site at http://digitalnewspapers.org, or you can search Google or Yahoo for "digital newspapers." We should be the first hit.

In 2001, the Marriott Library received a $93,000 Library Services and Technology Act (LSTA) grant to digitize thirty years of three weekly Utah newspapers. In December 2002, we launched the UDN Web site with ten thousand pages each from the *Wasatch Wave* (Heber City), the *Times Independent* (Moab), and the *Vernal Express*. This was our very humble beginning.

In 2003, the library received a second LSTA grant of $278,000 to continue digitizing newspapers. The library hired a project director and added 106,000 pages. Later in 2003, in a watershed event, the Institute of Museum and Library Services (IMLS) awarded the library a two-year, $1 million grant. With this grant we demonstrated that newspapers could be digitized and served over the Web to a wide audience on a large scale. We added more than 268,000 pages, distributed content to BYU and USU, and expanded coverage of Utah to include 27 of the 29 counties.

In 2004, the National Endowment for the Humanities (NEH), in conjunction with the Library of Congress, announced a national program to build a digital repository of historical newspapers, called the National Digital Newspaper Program (NDNP). In 2005, the Marriott Library, along with five other institutions across the country, received a two-year grant for NDNP Phase 1. With that grant we added 105,000 pages from 1900–1910. More recently, in 2007, the library received another two-year NDNP grant for Phase 2, through which we will process another 100,000 pages.

Along the way, several Utah public libraries, assisted by UDN, donated funds and/or obtained other LSTA grants to digitize their local newspapers. Just to name a few: in 2003 the Weber County Library provided $50,000 to digitize the *Ogden Standard* and its predecessors; between 2004 and 2006, the Park City Library and Historical Society raised $15,000 for the *Park Record*; and in 2007, Myton City with the Duchesne County Library raised $20,500 to digitize nine titles from Duchesne County. The Davis County Library established a relationship with the *Davis County Clipper* to fund the *Clipper*'s inclusion in UDN and continues to fund its digitization, narrowing the gap to the present day issues. In total, we have raised $2.5 million over the past five years. We have been and will remain an entirely soft-money program.

In the early days, we were among a small group of states actively building digital newspaper collections. Our friends at the University of Denver were experimenting via the Colorado Digitization Project (CDP), and the University of Florida was active as well.[1] At the same time we received our National Leadership Grant (NLG) from

---

[1] The Colorado Digitization Project has evolved into the Collaborative Digitization Program, which, not coincidentally, has the same CDP acronym.

IMLS, the CDP and the University of Richmond received similar National Leadership Grants from IMLS to launch/expand their collections.

Dissemination of our model was an important component of our IMLS grant. Accordingly, we presented our model to many groups across the country and encouraged them to initiate their own projects. Among the places we presented were the American Library Association conference in San Diego; the Center for Research Libraries in Chicago; the National Association of Government Archives and Records Administration, the Society of American Archivists, and the Council of State Archivists joint conference in Washington, DC; IMLS's WebWise conference in Los Angeles; digital newspaper forums in Denver, Colorado, and Portland, Oregon; statewide digital "summits" in Pierre, South Dakota, and Boise, Idaho; and numerous locations across Utah. We presented twice to the NEH and Library of Congress in Washington, D. C., demonstrating that large digital newspaper collections could be successful. Within six months of our second presentation, NEH announced the NDNP program. Today we are the only program to receive both an IMLS and NEH grant for newspaper digitization.

UDN has also emphasized training as an important part of our program. In July 2005, we hosted a three-day training program on how to launch a digital newspapers project. Twenty-three people attended from twelve different states and Trinidad and Tobago. Two future NDNP participants, the University of Kentucky and the University of California, were also in attendance.

We have won several awards as well, including an Award of Merit from the American Association for State and Local History and the John E. Jones Award from the Utah Press Association.

In May 2006, we were invited to host the annual conference of the Newspapers Section of the International Federation of Library Associations, as well as the annual meeting of the NDNP by the Library of Congress and NEH, respectively. We had a full week of intensive newspaper sessions at the University of Utah. One hundred people from ten countries attended the International Federation of Library Associations Conference, where noted speakers presented the latest ideas and developments. UDN was among the national and international programs that were featured.

At this writing (late 2007), UDN houses 570,000 newspaper pages, containing over five million articles. It holds fifty titles, including the first issues of the *(Salt Lake City) Deseret News* (1850) and the *Salt Lake Tribune* (1871), which today are the two leading newspapers in the state. The time period covered in the various papers is 1850–1969. We remain very active, with plans to expand the collection by 40 percent to more than 800,000 pages by the end of 2008.

Web site usage increased fourteen-fold from 60 visits per day in 2003 to 830 in 2006, when the program grew for eleven consecutive calendar quarters. It is amazingly popular with those who use it; 88 percent of readers rate us as "good" or "excellent."

Included in the collection you can find the *(Salt Lake City) Broad Ax*, an African American newspaper run by Julius Taylor around the turn of the twentieth century.

THE BROAD AX.

HEW TO THE LINE.

VOL. I.   SALT LAKE CITY, UTAH, AUGUST 31, 1895.   No. 1.

SALUTATORY.

With this edition we present to the good people of Utah, the initial number of the "Broad Ax." Launched upon the perilous sea of journalism, with a firm belief that it will be appreciated by those who will take the pains to peruse its columns, we desire to say, this journal will stand for the equal rights of all men before the law, for the interests of Salt Lake City, and the new State of Utah. It will be Democratic in politics, advocating the immortal principles of Jefferson and Jackson; it will stand for the

TO THE COLORED PEOPLE OF UTAH.

To the colored people of Utah we have a word to say.

The editor of this little journal, being himself a colored man, feels at liberty to address his people on terms of candor and equality, and therefore we ask you to come and "let us reason together."

There has been no race of people on earth, since the flight of time, who have been so much despised, downtrodden, and misused, by their fellow men as the negro race. Originating in the jungles of a savage continent, they have been brought to civilization through the

tionary war. It is true that the people of New England washed their hands of the sin, Pilate-like, by selling their slaves to their brethren of Virginia, Georgia and the Carolinas, because it did not pay to keep them so far north. It is true, many of them then turned abolitionists, and clamored for the liberty of the blacks, thus showing they were first actuated by a mercenary motive, and then, envious of the apparent prosperity of the southern planter, made a virtue of their jealousy to despoil their neighbors.

There is no logical reason why a colored man should be a Republican,

IN THE PUBLIC EYE.

Of all the names suggested as candidates for seats in the U. S. Senate, from the new State of Utah, there is a unanimity of sentiment among Democrats that the senatorial toga should rest upon the shoulders of Hon. J. L. Moses Thatcher of Logan. They are the logical candidates of the Democratic party. To Mr. Rawlins the people of Utah owe a debt of gratitude, for which a seat in the Senate would be but partial payment. As a champion of the interests of our people, and the work accomplished by him, while a dele-
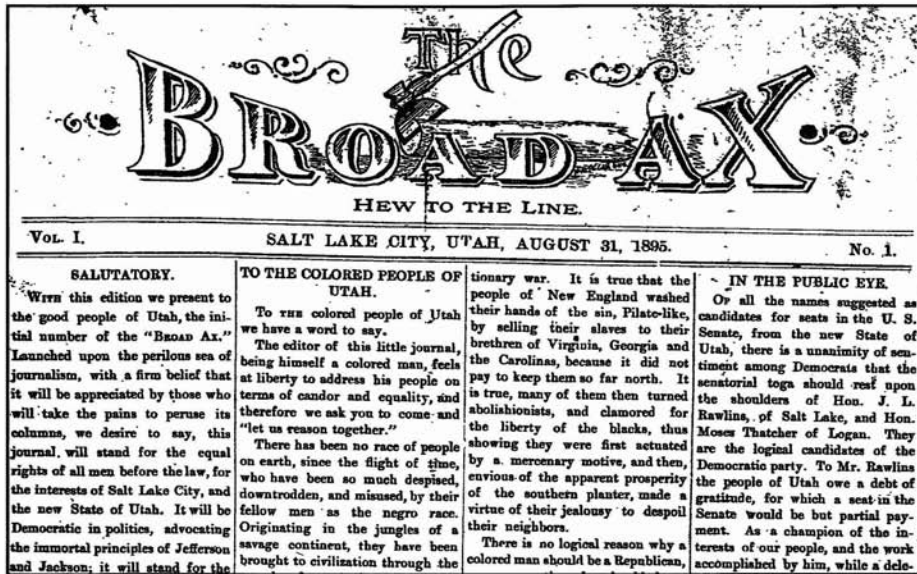
Figure 2. The first issue of the *Salt Lake City Broad Ax*, 31 August 1895. "Hew to the Line." This article can be found in the University of Utah's J. Willard Marriott Library's Utah Digital Newspapers Collection at http://udn.lib.utah.edu/u?/broadax,1.

(See Figure 2.) Taylor later moved the *Broad Ax* to Chicago, where he earned some national acclaim. Also included is the *Topaz Times*, the newspaper from the World War II Japanese internment camp near Delta, Utah. Significant misreportings are also noteworthy, as when the *Manti (Utah) Messenger* 19 April 1912, misreported that on the *Titanic* "Three Thousand Passengers Have Narrow Escape from Watery Grave." Important national events can be viewed through the small rural newspapers, as the *Vernal (Utah) Express*, 11 December 1941, declared, "Japan Attacks US" on page 3. This just happened to be their extravagant Christmas edition.

As more readers access UDN, they use it as a source for various types of research. Val Holley, who lives in Washington DC, recently published an article entitled "Leo Haefeli: Utah's Chameleon Journalist" in the Spring 2007 issue of the *Utah Historical Quarterly*. He listed forty references to articles in UDN. Roy Webb, staff member at the Marriott Library, presented on the history of Utah soccer at the 2007 Utah Historical Society annual meeting. His only source for its early twentieth-century history was UDN. Genealogists are our biggest group of readers. One of them, Thomas Billis, from West Jordan, found an article about himself from 1930, when he was five years old. An Australian academic researcher used UDN for researching the cement industry in 1908. Community theater guilds read advertisements to see popular fashions of the time periods in which their plays are set. Even commercial firms use our collection; a food industry representative used UDN to research the history of tofu and soy food production in the West. The interesting uses of UDN grow every day.

Besides containing huge volumes of Utah history from large cities and small towns, several papers carried national and international news. UDN contains a wealth of material well beyond Utah's borders. Of our readers, 42 percent live outside the state.

By providing newspapers online, accessing content is as simple as connecting to our Web site. The entire collection is full-text searchable through a search box on the home page. Likewise, individual newspaper titles can be searched from their own Web pages. There the user may browse by publication year, view all births, marriages, and deaths announced, and perform specialized searches.

One of the best and most popular ways to browse for regional news is our county map. The user can view a map of Utah and select available titles in each county.

As of this writing, keyword searches find 10,170 results for "suffrage," 8,422 results for "statehood," and, using an exact phrase search, 155 results for "Butch Cassidy." Using the advanced search feature, users may limit searches to a particular group of titles and specific fields. Searching for "prohibition" and either "women" or "ladies" in Davis County papers, produces 202 results.

With an estimated eight to ten million pages of newspaper content in the state to digitize, content selection is paramount. The more practical and technical considerations are availability, quality, and format of the source materials. From our own experience, we determined that the image quality scanned from original paper sources was superior to those from microfilm. Utah has a dry climate and a culture of safe-guarding historical artifacts, so we are often able to find high quality originals. Some are from library or historical society collections, but also, as in the case of the *Myton (Utah) Free Press,* they might be in an old family chest where they are kept for "safe" storage.

When microfilm is used, the master microfilm reels must be located and copied, because patron copies being used for research are often too scratched and worn. Processing from paper is slightly more expensive than microfilm, which can be a determining factor if available funding is tight. The current UDN collection contains a fairly even split between paper and microfilm.

Once we locate good source materials, the UDN advisory board reviews the titles. Our board is an extremely knowledgeable group of local historians, librarians, writers, and industry representatives. They identify important titles of the times, their editorial policies, and how they complement other titles already in the collection. The board also provides additional historical context which cannot be found in a normal catalog record. They then help prioritize titles to add to UDN. We also consider user requests. Once we add a portion of a county paper, local demand surges for more content from that county to be added either from a rival paper or the current title online.

Initially, UDN focused on rural weeklies to provide a broad representation across the state and digitize over a longer time period. Currently, we are focusing on larger urban papers with national interest stories as part of the NDNP. Because NDNP's scope is somewhat limited, we are seeking additional funds to digitize titles from rural counties and foreign language titles from early immigrant communities. As other libraries, organizations, and individuals are able to raise funds for us to digitize particular titles, we do so. We haven't had to turn down content yet!
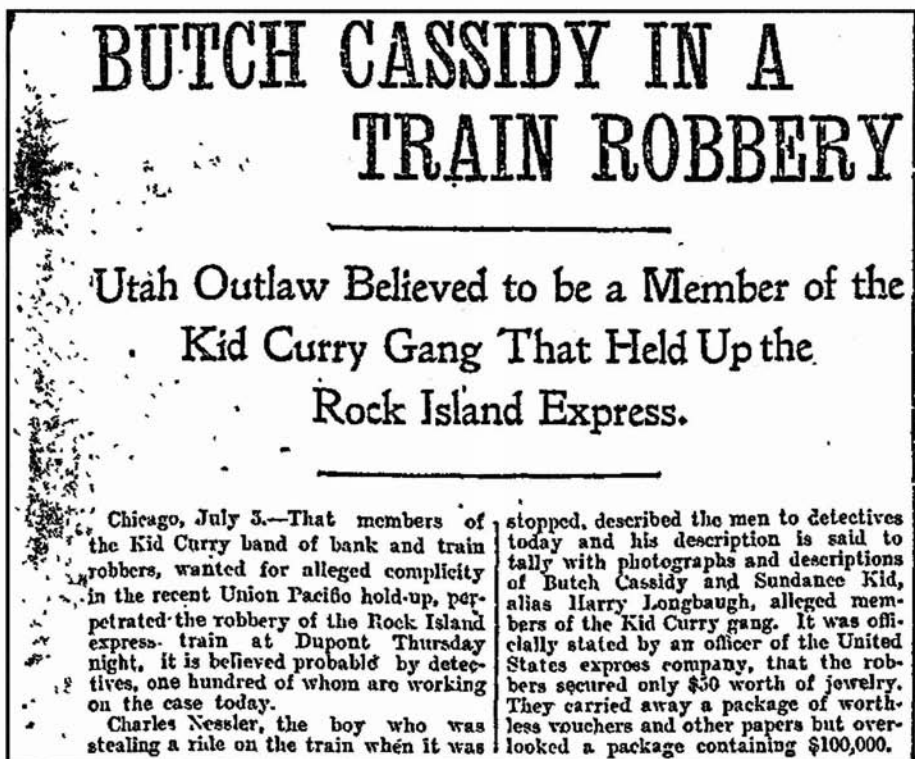
BUTCH CASSIDY IN A TRAIN ROBBERY

Utah Outlaw Believed to be a Member of the Kid Curry Gang That Held Up the Rock Island Express.

Chicago, July 5.—That members of the Kid Curry band of bank and train robbers, wanted for alleged complicity in the recent Union Pacific hold-up, perpetrated the robbery of the Rock Island express train at Dupont Thursday night, it is believed probable by detectives, one hundred of whom are working on the case today.

Charles Nessler, the boy who was stealing a ride on the train when it was stopped, described the men to detectives today and his description is said to tally with photographs and descriptions of Butch Cassidy and Sundance Kid, alias Harry Longbaugh, alleged members of the Kid Curry gang. It was officially stated by an officer of the United States express company, that the robbers secured only $50 worth of jewelry. They carried away a package of worthless vouchers and other papers but overlooked a package containing $100,000.

Figure 3. One of the many exploits of Butch Cassidy, related in the *Ogden (Utah) Standard*, 5 July 1902. This article can be found in the University of Utah's J. Willard Marriott Library's Utah Digital Newspapers Collection at http://udn.lib.utah.edu/u?/ogden7,71734.

The digitization process starts with a newspaper page, either an original or on microfilm. The first step creates 400-dpi, grayscale, digital images in TIFF format. Then software crops the images (to eliminate edges on the frame that are not part of the page), de-skews them (to orient the page directly up and down), and de-speckles them (to remove small blurs or dots picked up by the scanner).

Each page goes through an article "zoning" process where human beings identify and classify them as news, an advertisement, or birth, death, or marriage announcement.[2] An automated process performs Optical Character Recognition (OCR) against each article and creates a file of the article text. After generating the "raw" text, another automated process filters it through English dictionaries, a Utah place-names dictionary, and an extensive surnames list. The OCR-generated text is not 100 percent accurate. (It averages 70 percent, according to our own survey.) Still, it provides keyword access

---

[2] Some digitization processes use software to segment the newspaper page into article zones. This automated solution, however, has yet to prove to be reliable enough for us to incorporate into our process. We much prefer the higher quality and reliability that comes from our manual process.

HILLSTROM IS BURIED; FUNERAL TYPICAL ONE

CHICAGO, Nov. 24.—They're burying Joe Hillstrom tomorrow. It's a big event in Chicago's "rebel" ranks.

The ghetto, the slums, the lodging house quarters and the manufacturing districts are buzzing with preparations for the funeral of the Industrial Workers of the World song writer who was executed in Utah after conviction of murder. Anarchists, nihilists, sabotagists and secret organizations whose existence is based on opposition to present social and economic systems, are taking part in the honorary rites over the body of their dead comrade.

Handbills in a dozen different languages have been circulated, urging large attendance at the West Side Auditorium at 10 o'clock tomorrow morning. Today hundreds of persons passed into the undertaking rooms and viewed the body of the poet whose full name was Hillstrom.

An Industrial Workers of the World band of forty pieces will furnish funeral procession music. The Russian mandolin club will provide music during the services.· Hillstrom's own songs will be sung and played

Black bordered handbills tied with cord in the organization color of red and black have been circulated with the program for the exercises In black type on the front of one is the inscription:

"In memory of Joe Hill, murdered by the authorities of the state of Utah, November 19, 1915. He died that men might live. We never forget."

O. N. Hilton, attorney of Denver, who defended Hillstrom and who represented the defense in the Haywood, Pettibone and Moyer trials, will deliver the funeral oration, which will be followed by addresses by James Larkin of Dublin, the man who led the 1st dock workers' strike in England, and William D Haywood.

Bill Lorton, one of Hillstrom's companions, accompanied the body from Salt Lake City to Chicago.

Figure 4. *Eastern Utah Advocate*, 26 November 1915, reporting on the death and burial of labor activist Joe Hill. This article can be found in the University of Utah's J. Willard Marriott Library's Utah Digital Newspapers Collection at http://udn.lib.utah.edu/u?/easternutah,64929.

to the content that is impossible with microfilm. Two people separately transcribe the masthead and article headlines and subheadings. This insures that headings and subheadings are nearly 100 percent accurate.

Up to now, we have stored an image of each article in a separate file, as well as an image of the full page. These images are bi-tonal PDF files with embedded text. We use grayscale PDF files for papers if they require a higher level of quality and detail to read the text, such as the early *Salt Lake Tribune*, which has up to ten columns on a page.

As technology develops and more users switch to higher speed Internet connections, we will advance to higher quality and more user-friendly images. This presents a significant trade-off between image quality and speed of download. We are transitioning to grayscale images in the JPEG2000 file format, which allows the user further zooming capabilities, to "clip" a section of a page, and to highlight an article on the page image.

Original newspapers are returned to their owner after scanning is completed. Because of the fear that some may rely on our digital images as a preservation means and destroy the originals, the library has considered acting as a warehouse for all original newspapers in the state. However, our charter at UDN is strictly newspaper digitization. Consequently, we consider newspaper storage and preservation beyond the scope of our program and have deferred to the Marriott Library to lead this effort. At this time, while no systematic program for receiving print originals has been established, the library remains open to donations of original newspapers.

The University of Utah is one of eight awardees in Phase 2 of the NDNP. The program aspires to eventually include content from all states and territories hosted at "Chronicling America" (http://www.loc.gov/chroniclingamerica/). The scope of the project is English-language papers, primarily scanned from microfilm, from the time

period 1836–1922. For states that participated in the NEH-sponsored United States Newspaper Program (USNP), which cataloged and preserved newspapers on microfilm, this is a natural transition. States that did a majority of their microfilming during the USNP process generally benefit from more recent and higher quality microfilm images. Microfilm remains the preservation medium for NDNP, although we provide "preservation" quality, i.e., high resolution, digital images to NDNP as well. The catalog records created during USNP are available on the "Chronicling America" site, which is a great resource to track down where newspaper titles are available, whether in print, on microfilm, or digital form. The NDNP content we generate will be available on the UDN Web site, but Utah content digitized from other projects will not be available from Chronicling America. Other western state awardees include: California (http://cdnc.ucr.edu/), Minnesota, Nebraska, and Texas. Each of them will eventually have an institutional Web site for their content.

As the Utah Digital Newspaper program continues to develop in the future, it will incorporate technological innovations to aid in searching or browsing. The most frequent suggestion from our users is to "add more content," and we intend to oblige. We hope to broaden our user community to more researchers and to the K–12 audience. In order to provide continued long-term growth, our goal is to become financially self-sustaining.

Being exclusively a soft-money program means we constantly need to raise money to expand the collection. To date, that has amounted to an impressive $2.5 million, but our needs continue to exceed our resources. We are planning to continue with the NDNP and utilize NEH funding for as long as possible. However, that program will not fund content beyond 1922, so we will have to raise local funds for content from 1923 and beyond.

We have had tremendous success in our first five years, digitizing 5,000,000 pages. But with our long-term goal exceeding 8,000,000 pages, we have only scratched the surface. We still need many millions of additional dollars to reach our fund-raising goal and to begin to consider our collection complete.

The first newspaper issue published in Utah was volume one, number one of the *Deseret News* on 15 June 1850. On the front page, middle column is this paragraph:

> A paper that is worth printing, is worth preserving; if worth preserving, it is worth binding; for this purpose we issue in pamphlet form; and if every subscriber shall preserve each copy of the "News," and bind it at the close of the volume, their children's children may read the doings of their fathers, which otherwise might have been forgotten; ages to come.

These inspiring words were appropriate for newspaper publishing then and are just as applicable today for newspaper digitization. Internet access to keyword-searchable databases is revolutionizing how historic newspapers are utilized. We have before us the research tools to bring "the doings of our fathers" to life. Let us begin.

■