



Case Study ■ Library of Congress National Digital Newspaper Program

Chronicling America: Digitizing millions of pages of historic newspapers

Challenge

America's newspapers are an extraordinary historic treasure chronicling the narrative of American history, culture, and life itself. Across the country, these newspapers have been preserved at the state and national levels by local, state, and federal conservators – using microfilm.

While they exist on microfilm, most newspapers are not easily accessed and available to the public to utilize for research.

Unfortunately these windows into history are rarely used in their current form.

In 2004, the Library of Congress, in partnership with funding from the National Endowment for the Humanities (NEH), began the National Digital Newspaper Program (NDNP), through which the NEH awards states with grant funds to digitize their historic newspapers.

The NDNP embarked on a daring mission to digitize and make available newspapers chronicling American history across all 50 states. The work involved in creating searchable and viewable newspaper pages in a free online ecosystem covering a period starting just after the ratification of the Constitution, through the Civil War, America's great march west, the Industrial Revolution, World War I, and so much more. It is quite a large endeavor, and required a unique solution.

Apex's NDNP solution makes every newspaper's text digitally searchable, allowing microfilm to act as the preservation mechanism, and removes the format barrier between historic newspapers and the general public's access to them.

Chronicling America: Digitizing Millions of Pages of Historic Newspapers



The Solution

That's where Apex CoVantage and its proprietary Intelligent Zoning and Algorithmic Conversion (IZAAC) technology come in. In 2005, Apex began serving the Library of Virginia in the very first grant phase of the NDNP program. Today, Apex is honored to be the longest-serving NDNP solutions provider, having worked with more than a dozen state NDNP awardees to digitize millions of pages of newspapers.

To begin the digitization process, NDNP awardees select newspaper titles for digitization, create duplicate microfilm copies, and supply the derivative microfilm to Apex for processing. Apex then scans and transforms page images into a set of digital deliverables. To accomplish this work, Apex uses a combination of human intellect and IZAAC's sophisticated signal processing technologies to disaggregate newspaper issues and pages into granular elements, such as the headline, subtitle, images, and captions. A series of sophisticated tasks are performed using both human and machine processes to then re-aggregate digital newspaper pages.

Monthly, Apex delivers multiple NDNP awardees with tens of thousands of digitized newspaper pages. Files are then submitted by the awardee programs to the Library of Congress, which evaluates the digital files for technical quality and accuracy. Digital files include TIFF, JPEG 2000, PDF, and METS/ALTO XML deliverables in accordance with NDNP technical specifications.

IZAAC technology enables the inclusion of crucial metadata about the digitized pages including title, date, issue, Library of Congress Catalog Number, and more, allowing librarians richer insight into their digital collections. But as advanced a technology solution as IZAAC is, Apex also incorporates human eyes into the process, checking to ensure all information has been identified accurately.

"Library specifications are usually different from the private sector, and a lot more stringent," explains Errol Somay, Director of the Virginia Newspaper Project, managed by the Library of Virginia. "To get a vendor to understand that can sometimes be a little difficult. They say, 'Oh yeah, we can do that,' and then you find out, no they can't. Because they're used to a different standard."

That's never been a problem with Apex, however, Somay said.

Once the digitized pages are accepted by the awardee and approved by the Library of Congress, they're ready to be uploaded to Chronicling America, the massive online database where all the content is made accessible to the public.



Results

Apex's NDNP solution makes every newspaper's text digitally searchable, allowing microfilm to act as the preservation mechanism, and removes the format barrier between historic newspapers and the general public's access to them.

Often, the first six months of each NDNP grant cycle is spent hiring staff, choosing a digitization partner, and determining which titles to digitize. That leaves even less time for all 100,000 pages allowed under the program to be digitized in an 18-to-24-month window.

Apex has an established reputation for meeting all of its obligations on deadline, with minimal rework required. It also provides awardees with access to its Basecamp project management instance, ensuring total transparency while allowing stakeholders insight into the status of project workflow.

Through Apex's involvement from the outset of the NDNP program, it often serves as a vital consultative resource to librarians and project managers new to the process, answering questions and helping ramp-up new programs together.





Content & Media Solutions

About Apex CoVantage

Apex is a leading supplier of data conversion, editorial, and content-enhancement services to publishers worldwide. By balancing leading technology and broad industry expertise, Apex delivers exceptional content and media solutions to its customers. Apex has been managing large-scale digitization projects for libraries, publishers, and media companies around the world for nearly 30 years.

- +1.703.709.3000
- content@apexcovantage.com
- @apex_content
- linkedin.com/company/apex-covantage
- facebook.com/ApexCoVantage

apexcovantage.com

Apex CoVantage
 200 Presidents Plaza
 198 Van Buren Street
 Herndon, VA USA 20170-5338

"We're dealing with newspapers, and when it comes to digitizing newspapers, you have a couple of real big problems: the size and format, the columns, which creates problems for text-readability, and the sheer number of pages. So what I'm looking for is someone who actually can do this for me according to our specs, and at a competitive price."

-Errol Somay,
 Director of the Library of Virginia's Virginia Newspaper Project