



## CASE STUDY - NEWSPAPER DIGITIZATION

# Yale Daily News

Digital Divide Data and the Yale University Library partner to digitize Yale Daily News content spanning more than 120 years.

### Business Challenge

The Yale Daily News (YDN) is the oldest college daily newspaper in the United States. Many of its student editors, writers, and contributors have gone on to prominent careers in journalism and public service. To preserve this important legacy and to make the YDN available to the widest possible audience, the Yale University Library engaged Digital Divide Data to begin digitization of the entire backfile of the newspaper – from 1878 through 2000. The university had very specific goals for the project: to present the reader with a full page view of the original newspaper; provide the ability to zoom in on details and select articles or photographs on a page; conduct a word, name, or phrase search across all digitized YDN issues; and print out or download copies of articles and pages. Yale required DDD to meet the Library of Congress' METS/ALTO standards for newspaper digitization, archival digital imaging and XML output.

### DDD Solution

To meet the challenge of digitizing 123 years' worth of YDN content, DDD worked closely with technology partner Brechin Imaging. Together, we developed a unique scanning solution for large format newspapers using a custom-built scanning rig and artful reprocessing. This approach helped to achieve easy online reading and improved OCR accuracy. Brechin was also able to integrate scans from both paper and microfilm where pages were missing from paper copies. A second technology partner, Content Conversion Specialists, provided their docWorks software for METS/ALTO conversion. Word positions, article boundaries, document structure, and OCR output were encoded by DDD using docWorks. Cleanup work, such as verification of article boundaries and quality control checks, were performed by DDD staff in Cambodia. DDD shipped all files and digital assets to the university ready for import into Yale's digital library system.

“Working with Digital Divide Data has been an absolute pleasure! The Yale Daily News certainly stands out as one of our most successful digitization projects this year.”

- Frederick Martz, Office of Digital Assets and Infrastructure, Yale University Library

### Impact

The YDN archive project is supported by The Oldest College Daily Foundation, the Yale University Library, and the class of 1945W, as well as other classes and individual donors. In Phase One of the project, approximately 25,000 pages were digitized out of a total of some 135,000 pages. This phase of the YDN online collection was officially launched during Yale's 2008 reunion weekends. At these reunions, the project team hosted a kiosk showcasing the online archive. Interacting with the archive, alumni found vivid memories of their time at Yale, such as the 1941 Yale-Princeton football game, and citations of parents and grandparents. The children at the reunion especially loved seeing photos of their parents. Issues of the YDN from important periods in Yale's history are now available online and searchable through the University Library's Yale Daily News Historical Archive. Additional years will be added to the online collection as funding becomes available. The YDN is an important resource in the Library's department of Manuscripts and Archives and is heavily used by researchers, alumni, students, and staff. A fully indexed online version will dramatically enhance access to the YDN both locally and globally.

Digital Divide Data powers the data entry, XML conversion and digital preservation needs of publishers, libraries, content hosts, academic researchers and businesses world-wide. We deliver these services on the foundation of a socially responsible model that creates opportunity for the world's poorest citizens to earn competitive wages, complete their education, and achieve upward mobility through working in our business.

#### DIGITAL DIVIDE DATA

WEB [www.digitaldividedata.org](http://www.digitaldividedata.org)  
EMAIL [sales@digitaldividedata.org](mailto:sales@digitaldividedata.org)  
PHONE +1.212.461.3700

Find more about our services here: <http://www.digitaldividedata.org/services/newspapers/>

© 2010, Digital Divide Data. All rights reserved. All trademarks and registered trademarks are the property of their respective owners.