The CDNC: An Introduction

Brian Geiger, Director
Wednesday, December 17, 2014
12 Noon

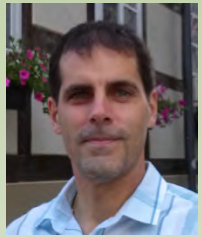
Poll

California Digital Newspaper Collection

- ≈ 1,000,000 pages
- ≈ 3,000,000 embargoed pages
- 1846-present
- Available to everyone at cdnc.ucr.edu
- Partnerships with local, state and national institutions

CDNC
California Digital Newspaper Collection
California Digital Newspaper Collection
A Freely Accessible Repository of Digitized California Newspapers from 1846 to the Present

Who We Are



Brian Geiger
Director

Christine Straitt
Newspaper Cataloger

Luis Baquera
IT Administrator

Jay Yasul
Admin Assistant

Ginger Schilling
Data Analyst

Center for Bibliographical Studies and Research

UC RIVERSIDE UNIVERSITY OF CALIFORNIA

CDNC California Digital Newspaper Collection

California Digital Newspaper Collection

A Freely Accessible Repository of Digitized California Newspapers from 1846 to the Present

Related Projects

- Bibliography and union catalog of newspapers in CA institutions
- Holdings abbreviations:
  - "OR" = original newsprint
  - "FM" = positive microfilm
  - "FMM" = master negative microfilm
- "California Newspaper Microfilm Archive" holdings = CNMA film for digitization
- cnp.ucr.edu

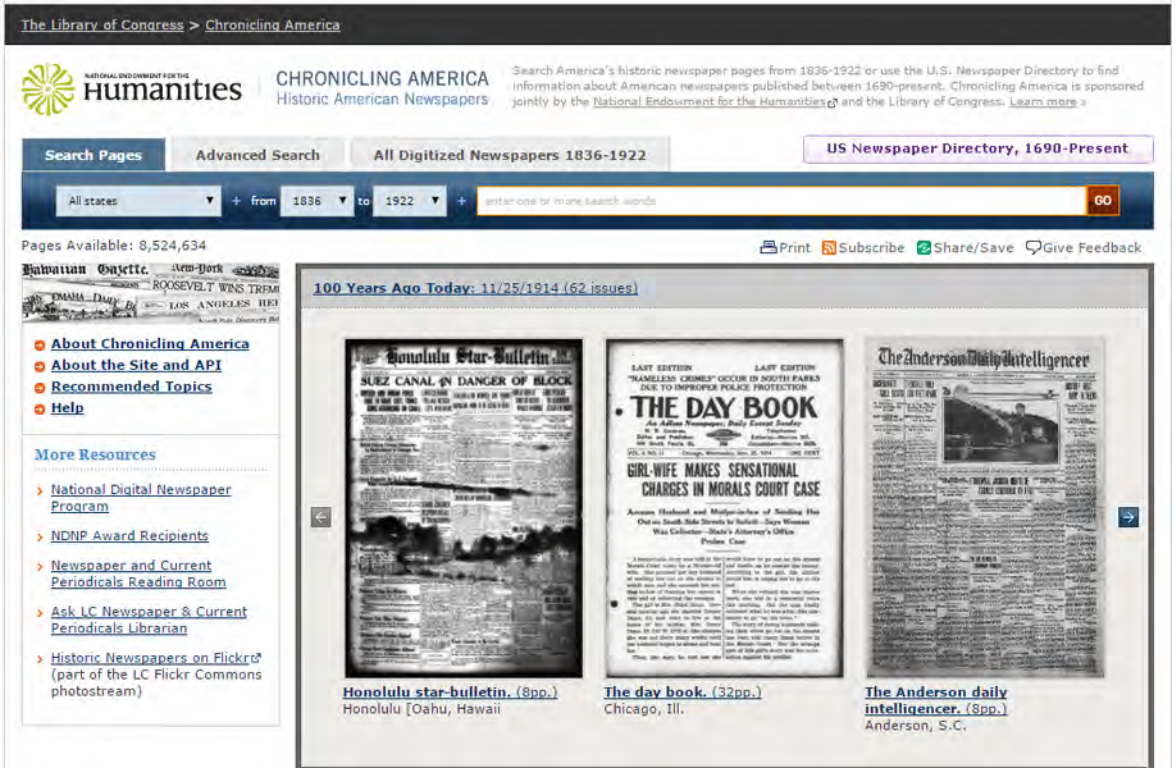
- Master negative microfilm archive
- 100,000 reels
- Stored at SRLF and NRLF
- Available for digitization
- cnma.ucr.edu

Agenda

- History and Background
- Standards
- CDNC Features
- Partnerships
  - California State Library/LSTA
  - Local Institutions
  - Ancestry.com
  - Born Digital Project

2005-2011: National Digital Newspaper Program (NDNP)

- Project of the Library of Congress and National Endowment for the Humanities
- Participating states send content to LC
- Content available at chroniclingamerica.loc.gov
- CDNC produced 300,000 pages, available at cdnc.ucr.edu and ChronAm



2007-present: LSTA/CSL

- Library Services and Technology Act (LSTA) grants
- LSTA administered by the California State Librarian
- Produced over 450,000 pages
- CDNC officially launched 2007

CALIFORNIA STATE LIBRARY

INSTITUTE of Museum and Library SERVICES

2008-present: Local Partnerships

- Produced over 200,000 pages
- Ancestry.com Collaboration
- Partnering institutions include
  - Sausalito Public Library
  - Coronado Public Library
  - Santa Monica Community College
  - Palm Springs Public Library
  - Tehama Public Library
  - Occidental College
  - Healdsburg Museum & Historical Society
  - San Bernardino County Historical Archives
  - Port of Los Angeles
  - Barbro Osher Pro Suecia Foundation

Overview

- Follow standards set by LC and NEH for NDNP
- Used around the world, including Australia, Europe and Asia

Details

- Scan from CNMA master negative film, occasionally newsprint or positive film
- TIFF archival image for every page
- Derivative files provide access

Why

- Digital preservation copies of newspaper
- Provide superior access than formats like PDF
- Usable across platforms
- Eliminate need to reprocess in the future

Standards

Scanning Vendor
TIFF Image per Page
Processing software
docWorks
Derivative Files
1. TIFF image
2. JP2 image
3. PDF page
4. PDF issue
5. METS/ALTO XML
Partnering Institution

Derivative Files

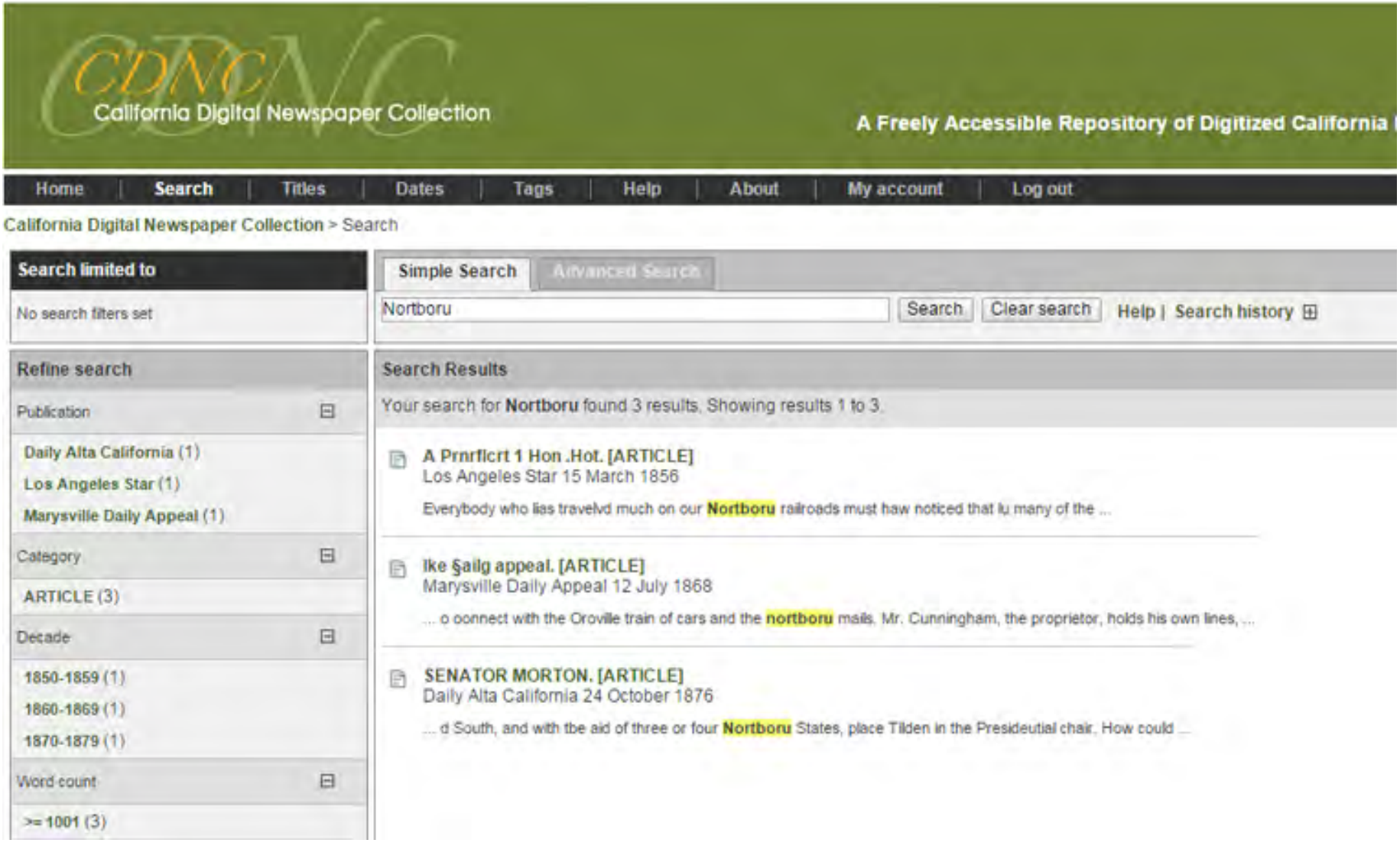
TIFF Images
- Digital preservation copy of page
- Soon downloadable for high-res reproductions

JP2 Images
- Used for page display in CDNC
- Available for high-res reproduction

PDFs
- Downloadable for low-res purposes

METS/ALTO XML
- METS defines the physical layout of the issue and pages
- ALTO defines page contents and computer-generated text (OCR)
- Allows sophisticated access and retrieval of data

CDNC Features



CDNC Features

CDNC California Digital Newspaper Collection
California Digital Newspaper Collection
A Freely Accessible Repository of Digitized California Newspapers from 1846 to the Present

CDNC Features

1.	Wes Keat	838,881
2.	annh	334,401
3.	rstew160	150,000
4.	Vince Miller	115,173
5.	Roger	110,023
6.	RIddings	99,999
7.	wmartin46	97,407
8.	Sharonvdc	96,792
9.	Conovaloff	73,186
10.	JFZ	70,166
11.	Mike Detwiler	63,786
12.	JeffD-OCRtf	57,565
13.	Toby	57,441
14.	mjbarkl	49,030
15.	Phyllis	40,855
16.	CathyMK	36,647
17.	Teresa Nemeth	35,268
18.	Sheila S	34,562
19.	Kevin Hopkins	31,319
20.	DeborahC	31,039
21.	nrowlett	30,408
22.	rieck	27,616
23.	JuneauJim	26,914
24.	kost	26,104
25.	stepsix	25,755
26.	astrayelmgod	25,152
27.	J Sergneri	22,691
28.	cferoben	21,458
29.	daveg	20,786
30.	Peter H	20,221

CDNC Features

Future Additions

- Automatically download high res TIFFs and JP2s
- Search embargoed content
- Robust user accounts
  - Save search histories
  - Save and organize search results
  - Take notes on search results
- Tag content with authorized terms like “Illustration,” “Birth/Death Notice,” “Ad”
- View titles by county

Partnerships

California State Library/LSTA:
5-Year Plan

Partnerships

Local Institutions

Funding

- CSL/LSTA 5-Year Plan
- CSL/LSTA Pitch an Idea: Local projects that can include newspapers, but…
- Local Funding Agencies: foundations, businesses, etc.
- Local Resources: fund-raising campaigns, donor, acquisition funds, etc.

How it Works

- CNMA master neg microfilm, local positives if necessary, occasionally newsprint
- CDNC bundles projects for lower pricing
- 2015 per page pricing: $0.06 scanning + $0.34 digitization = $0.40 total
- CDNC pays vendors and invoices institution for reimbursement
- CDNC QCs data then loads into live site and sends copies to institutions
- There is no charge to local institution for hosting data in the CDNC, but…

Partnerships

Local Institutions: Branded Website at CDNC

Partnerships

Ancestry.com

Overview

- Ancestry digitizes film at their cost
- Adhere to CDNC digitization standards
- Provide copies of all data to CDNC
- Available immediately at newspapers.com for a fee
- After 3-year embargo available for free to everyone at CDNC
- In 2015 hope to make embargoed data searchable but not viewable in CDNC
- Interested primarily in long runs, late 19th to early 21st century
- To date have digitized San Bernardino Sun, Santa Cruz Sentinel, & Oakland Tribune

Local Institution

- Clears copyright by researching or working with copyright holder
- Gets free access to titles at newspapers.com during embargo period

Partnerships

Born Digital Project

- Many newspapers no longer microfilmed
- PDFs can be preserved instead
- CBSR began collecting PDFs in 2010
- Publisher uploads PDFs as sending to printer
- CBSR developed special uploading software
  - cdnc.ucr.edu/pdfuploader
- Periodically process PDFs to include in CDNC

THE PENINSULA BEACON

SAN DIEGO COMMUNITY NEWSPAPER GROUP — THURSDAY, FEBRUARY 25, 2010

School board tries to close budget gap

By SEBASTIAN RUIZ | THE BEACON

The San Diego Unified School District Board of Education voted to cut about $63 million from the district's budget Tuesday by reorganizing central office departments, changing school programs, consolidating high schools with the fewest students and charging for student AP exams.

The board also voted to approve "negotiable" items to be approved by unions. These affect workers' and teachers' pocketbooks with five-day furloughs, raising the cost of medical benefits and supporting a 6 percent across-the-board salary rollback.

This most recent round of cuts also left approximately 230 full-time equivalent positions in limbo as the board worked to close a projected $79 million budget deficit for next school year.

Board of Education President Richard Barrera said tens of millions of dollars in budget cuts from the state over the last two years have put the board in a difficult position.

"This is our second round of going through [budget cuts]. It's very painful," Barrera said after the Feb. 23 meeting.

While the board tried to keep as much of the pain away from the classroom as possible, students stand to lose educational resources and programs.

The Off-Campus Integrated Learning Experience (OCILE) program takes students to Old Town for a week to teach them about cultural and racial diversity and history. That program was reduced to two days a week.

Friction on th

Young transient sees d sides of life in OB's trav

By KEVIN McKAY | PENINSULA BEACON

Editor's note: A new homeless issue has emerged in the community over the last couple of months in the way of young transients and behavioral problems in Ocean Beach. The issue has seemingly divided the community between sympathizers and those bemoaning an atmosphere of fear and intimidation near the sea wall. The Peninsula Beacon attempted to speak with several of the young nomads for insight. One wayward traveler was willing to share his views on the record. Here is his take from the youth perspective.

Lucas Swafford tries hard to dole out respect to the locals of Ocean Beach. He just wants to play his music, relax in the sun and get some of that same respect in return. Swafford, 25, is painfully aware of how he and some of his young transient friends are being perceived by the community.

Swafford is part of an ever-evolving enclave of young adults which has ebbed and flowed near the oceanfront over the last few months, calling themselves "urban travelers," camping out on the grass area of Veteran's Park, in the alcoves of businesses along Newport

Lucas Swafford, 25, said he has seen the good, bad and the ugly on the streets of Ocean Beach. A part of a self-named group known as the "urban travelers," Swafford said the vast majority of the 100 or so transient youths respect the local community and police themselves against the bad seeds. JW GRANT | THE BEACON

OBTC slates dialogue on

By KEVIN McKAY | PENINSULA BEACON

The Ocean Beach Town Council (OBTC) was scheduled to host a community dialogue last night, Feb. 24, on the recent influx of a band of young, aggressive panhandlers to the Newport Avenue, Ocean Beach Pier and sea wall areas in recent months.

The session was generally intended to address homelessness issues locally — but was expected largely to generate comments from all sides

How You Can Help

- Publishers stop uploading PDFs
- Local institution has access to account
- Check account and remind publisher to upload

Resources

Microfilm

- California Newspaper Project, http://cnp.ucr.edu: bibliography and union catalog of CA newspapers around the state. "FM" holdings are positive reels; "FMM" master negs; "OR" newsprint. Holdings marked "California Newspaper Microfilm Archive" are FMM managed by the CDNC and available for digitization or duplication.
- California Newspaper Microfilm Archive, http://cnma.ucr.edu: reel-by-reel database of master neg film available for digitization or duplication. Note not all reels have been entered, and it is best to double check availability in the CNP above.

Best Practices

- METS Standard: maintained by LC at http://www.loc.gov/standards/mets/
- ALTO Standard: maintained by LC at http://www.loc.gov/standards/alto/
- Historic Newspaper Digitization: a number of documents created by the CDNC, in part with LSTA funding, available at http://cdnc.ucr.edu/site/related.html

Resources

Documentation

- Info on “Born Digital” project and PDF uploader http://cdnc.ucr.edu/site/uploading.html
- Copyright http://cdnc.ucr.edu/site/USNewspapersandCopyright.html
- California counties 5-year plan http://cdnc.ucr.edu/site/cacounties.html#

Thank You!

Brian Geiger, Director
bgeiger@ucr.edu
951-827-7007
cbsr.ucr.edu

Infopeople

helping libraries think differently

Infopeople webinars are supported in part by the U.S. Institute of Museum and Library Services under the provisions of the Library Services and Technology Act, administered in California by the State Librarian. This material is licensed under a Creative Commons 3.0 Share & Share-Alike license. Use of this material should credit the author and funding source.

CDNC California Digital Newspaper Collection

California Digital Newspaper Collection

A Freely Accessible Repository of Digitized California Newspapers from 1846 to the Present