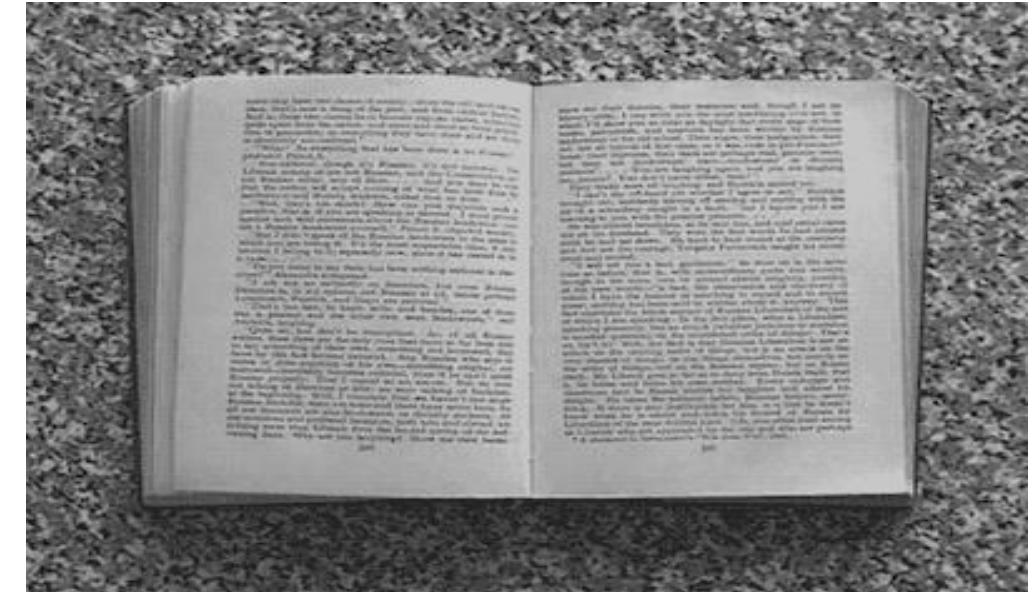
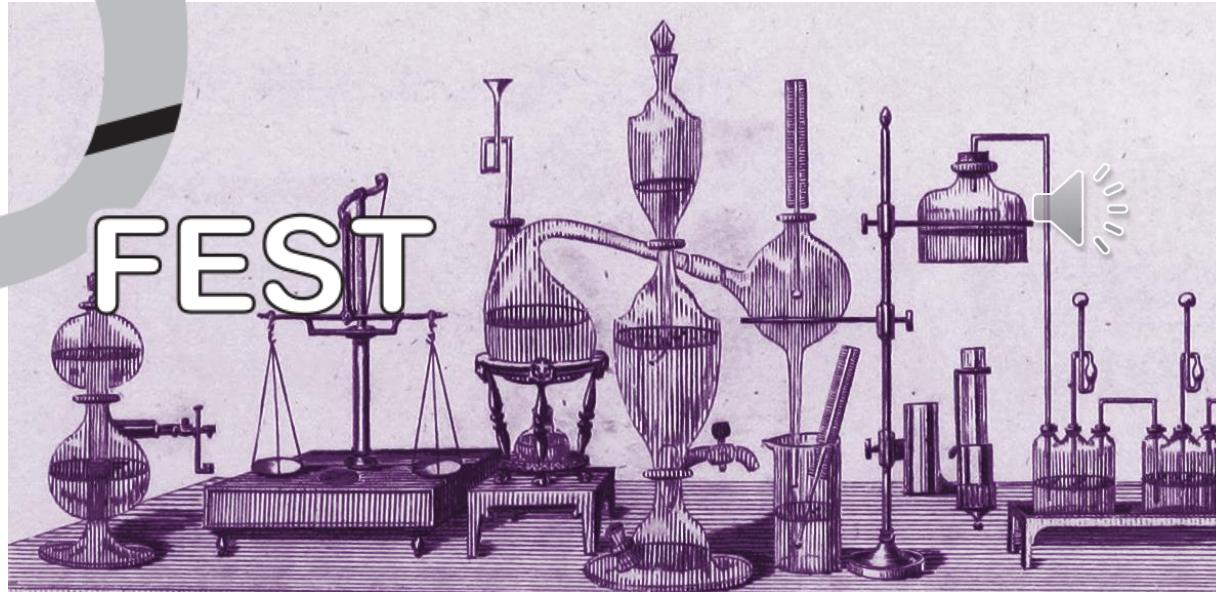


Digitalno objedinjavanje kroz razvoj „zbirki kao podataka“ i baštinskih suradničkih inovacijskih laboratorija



Dr. sc. Ivana Dević

Sveučilišna knjižnica u Splitu

Deseti festival hrvatskih digitalizacijskih projekata / NSK / 6. i 7. svibnja 2021.

AKM LABORATORIJI

- suradnička mjesta za znanstvena i interdisciplinarna istraživanja i za rad s korisnicima koji su zainteresirani za **korištenje digitalnih zbirki na eksperimentalan i inovativan način**
- Online ili u fizičkom prostoru
- Osiguravaju pristup digitalnim/digitaliziranim **zbirkama na razini podataka**
- **Korisnici:** istraživači, umjetnici, nastavnici i studenti, arhivisti, poduzetnici, programeri softvera i naravno knjižničari...
- **Suradnja:** podatkovna znanost, digitalna humanistika, računalne društvene znanosti itd.
- **Koji su ciljevi laba?:** pristup velikim paketima podataka digitalnih zbirki, računalno potpomognuto istraživanje, vizualizacija, analiza i upotreba podataka u drugim aplikacijama, STEM ili humanistički radni zadaci i projekti, ponovna upotreba digitalnih materijal na novi način

PREPORUKE LIBRARY OF CONGRESS LAB

- Pomak prema digitalnim istraživačkim laboratorijima – potencijalni istraživački projekti
- Laboratorij bi trebao biti uslužno orientiran na vanjske korisnike te ih podržavati u rastu i razvoju njihovih istraživanja
- Laboratorij treba biti obrazovni prostor
- Pristup zbirkama (digitaliziranim i digitalnim) na razini podataka i osigurava uključivanje zbirki u istraživačke tijekove rada na korisnički pristupačan način
- prilagoditi digitalizirane zbirke za korištenje digitalnih humanističkih metoda istraživanja, poput rudarenja teksta i podataka
- Transformacija obavljanja usluga od strane knjižničara u laboratoriju: identificiranje dostupnih podataka, rad s tim podacima za isporuku određenog korpusa, utvrđivanje porijekla podataka i pomoći u odgovoru na istraživačka pitanja s podacima
- Usluge laboratorija postaju dio uobičajenih usluga knjižnice

LIBRARY OF CONGRESS LAB

Library of Congress Digital Scholars Lab Pilot Project Report

The Lab envisioned by the authors in this report is based on successful models that support the development of pilot projects which have been drawn from interviews with experts and scholars and shaped by practical experience. The recommendations made in this report consider the Lab as part of the unique structure and offerings of the Library of Congress.

Michelle Gallinger and Daniel Chudnov
12/21/16

Gallinger, M. & Chudnov, D. (2016):
Library of CongressLab. Library of
CongressDigital ScholarsLab Pilot Project
Report. Dostupno na URL:
https://labs.loc.gov/static/portals/labs/meta/images/DChudnov-MGallinger_LCLabReport.pdf

International GLAM Labs Community

<https://glamlabs.io/>



British Library Labs – osnovan 2013, 150 skupova podataka (od 06/09/2020)

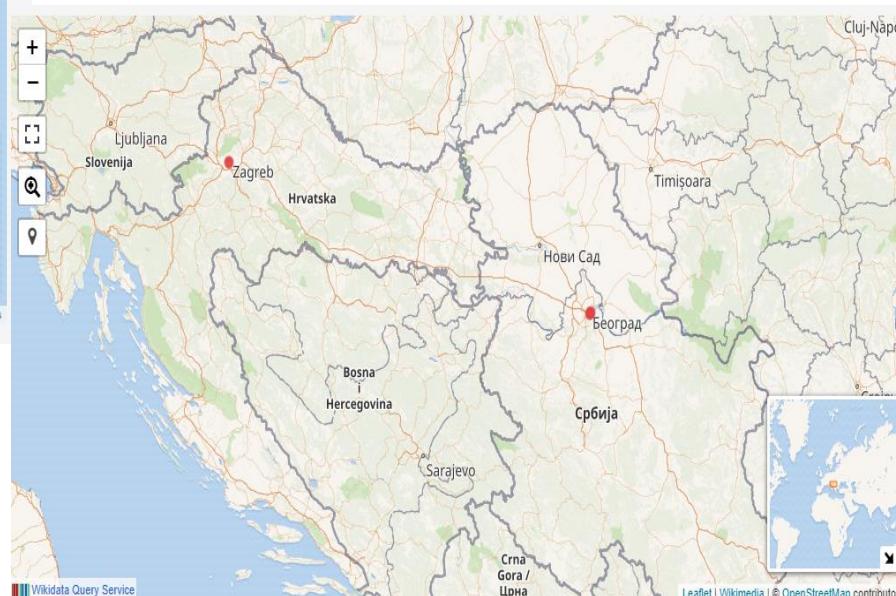
Dutch KB Lab – osnovan 2014,

Library of Congress Labs – osnovan 2017

Europeana Labs - 2014

Danish KB Lab – 2016

Austrian National Library (ÖNB-LAB) - 2018



NSK – otvoren 2020,

GlagoLab – Centar za istraživanje glagoljaštva Sveučilišta u Zadru - Senat Sveučilišta u Zadru na svojoj 5. Sjednici 2020,

Digitalni laboratorij – Institut za povijest umjetnosti



Zajednica svjetskih inovativnih kulturnih i istraživačkih institucija

Building Library Labs Survey 2018

Growing 'Library' Labs International Community

Podaci dostuni na URL: <https://goo.gl/66icov>

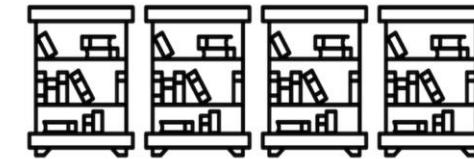
ZBIRKE KAO PODACI

- Bez podataka nema laboratorija, stoga digitalne zbirke trebaju biti dostupne u formi podataka - temeljnoj aktivnosti laboratorija
- Transformacija zbirki u „zbirke kao podatke“: identificirati zbirke i procijeniti njihovu prikladnost za laboratorije, opisati ih, učiniti ih dostupnim i ponovo upotrebljivim – > zbirke postaju djeljive kao podaci
- *Always Already Computational: Collections as Data* imala je za cilj: pronaći način dokumentiranja, razmjene iskustva i dijeljenja znanja kako bi „podržala korisnike koji žele raditi sa zbirkama kao podacima“ (Padilla, 2019)
- Ustanove kulturne baštine su rijetko izgrađivale digitalne zbirke ili stvarale pristup s ciljem da podrže njihovu računalnu upotrebu, osim iznimki (LC, Nacionalna knjižnica Nizozemske, British Library), (The Santa Barbara Statement on Collections as Data, 2017)
- Zbirka kao paradigma podataka nastoji potaknuti prošireni niz istraživačkog, pedagoškog i umjetničkog potencijala koji se temelji na računalnoj uporabi zbirki kulturne baštine. Zbirke kao podaci postavljaju pitanje što bi moglo značiti tretirati digitalizirane i izvorne digitalne zbirke kao podatke, a ne kao jednostavne surrogate fizičkih objekata ili statične prikaze digitalnog iskustva (Padilla, Thomas. “Collections as Data: Implications for Enclosure 2018.” *College & Research Libraries*, vol. 79, no. 6, 2018, <http://crln.acrl.org/index.php/crlnews/article/view/17003/18751>.)
- Korisnici sve više generiraju vlastite podatke i eksperimentiraju zajedno s AKM labovima za zajedničko generiranje novih skupova podataka
- Pristup zbirkama i velikim paketima otvorenih podataka i metapodataka povezanih s digitaliziranim i izvornom digitalnim zbirkama u svrhu njihove upotrebe na nove načine

ALWAYS ALREADY
COMPUTATIONAL - COLLECTIONS
AS DATA

HOME TEAM PARTNERS EVENTS RESOURCES UPDATES

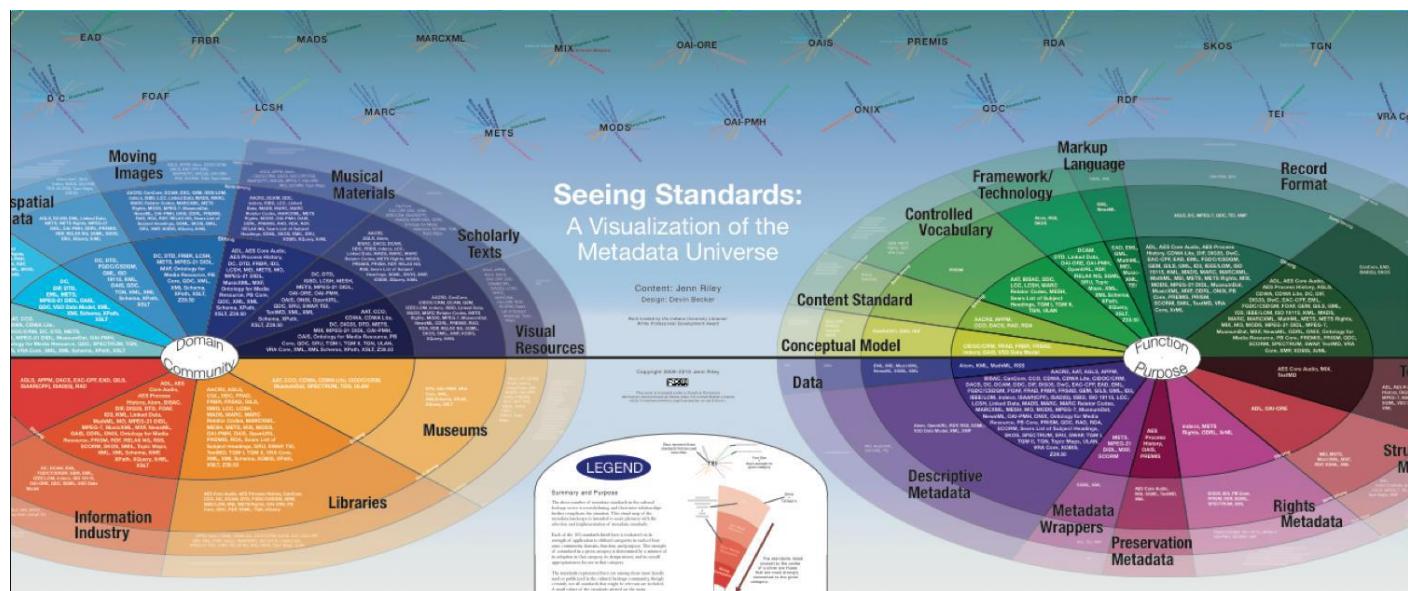
Always Already Computational



Collections as Data is an Institute of Museum and Library Services supported effort that aims to foster a strategic approach to developing, describing, providing access to, and employing new types of collections to support computationally-driven research and teaching in areas including but not limited to Digital Humanities, Public History, Digital History, data driven Journalism, Digital Social Science, and Digital Art History.

collectionsasdata.github.io

ANALOGNE ZBIRKE



Prevladavajući razvoj digitalne zbirke usredotočen je na preslikavanje tradicionalnih načina interakcije s objektima u digitalnom prostoru. Ovaj pristup ne udovoljava potrebama istraživača, studenta, novinara i ostalih koji bi željeli [raditi sa] računalnim metodama i alatima za tretiranje ... zbirki kao podataka. (Thomas Padilla, 2016)

PRISTUP : MOGUĆNOST PREUZIMANJA

MARC Distribution Services (data set)

[Print](#) [Subscribe](#) [Share/Save](#) [Give Feedback](#)

[« Back to Product Categories](#)

These electronic services are your source for MARC records, delivered via FTP (File Transfer Protocol). MDS offers cataloging records, available in original scripts, for materials in hundreds of languages, from Arabic to Yiddish, along with other non-roman alphabet languages such as Chinese, Hebrew, Japanese, and Korean. MDS provides national level cataloging for almost all English-language imprints. Each service is available as a current-year (January–December) subscription or a retrospective file.

[2019 MARC Distribution Service Record Counts and Prices](#)



[Read more about MDS >](#)

- > [Books All \(database\)](#)
- > [Books English \(database\)](#)
- > [Books U.S. \(database\)](#)
- > [Classification \(database\)](#)
- > [Complete Service \(database\)](#)
- > [Computer Files \(database\)](#)
- > [Copyright Cataloging: Monographs, Documents, and Serials \(database\)](#)
- > [Copyright Select \(database\)](#)
- > [Maps \(database\)](#)
- > [Music \(database\)](#)
- > [Name Authorities \(database\)](#)
- > [Serials \(database\)](#)
- > [Subject Authorities \(database\)](#)
- > [Visual Materials \(database\)](#)

MARC Open-Access

In addition to our MDS subscription service, we now offer an "open-access" provision, MDSConnect, at no cost to users. Our open-access service includes nearly 25 million MARC records, as distributed in the unabridged 2016 Retrospective file sets. These MDS record sets have been made available primarily for research and development usage. Records are available in two file formats - UTF8 and XML.

To explore this open data, please select from the links below. Within each link, you will be given the choice of selecting the file type (UTF8 or XML).

Please note that record sets are contained in zipped files and you may need to download several files.

- > [Books All](#)
- > [Classification](#)
- > [Computer Files](#)
- > [Maps](#)
- > [Music](#)
- > [Name Authorities](#)
- > [Serials](#)
- > [Subject Authorities](#)
- > [Visual Materials](#)

[FAQs](#)



[Getting Started](#)



[CHRONICLING AMERICA](#)
Historic American Newspapers

Search America's historic newspaper pages from 1777-1963 or use the U.S. Newspaper Directory to find information about American newspapers published between 1850-present. Chronicling America is sponsored jointly by the National Endowment for the Humanities and the Library of Congress. [Learn more >](#)

Pages Available: 18,001,700

[Print](#) [Subscribe](#) [Share/Save](#) [Give Feedback](#)

OCR Data

What is OCR?

Optical character recognition (OCR) is a fully automated process that converts the visual image of numbers and letters into computer-readable numbers and letters. Computer software can then search the OCR-generated text for words, phrases, numbers, or other characters. However, OCR is not 100 percent accurate, and, particularly if the original item has extraneous markings on the page, unusual text styles, or very small fonts, the searchable text OCR generates will contain errors that cannot be corrected by automated means.

Although errors in the process are unavoidable, OCR is still a powerful tool for making text-based items accessible to searching. For example, important concept words often appear more than once within an article. Therefore, if OCR misreads one instance of a key word in a passage, but correctly reads the second instance, the passage will still be found in a full-text search.

To enable research and external services Chronicling America provides bulk access to its OCR data. The table below itemizes a list of data files available for download. Each file will decompress into directory structure that lets you easily map the OCR file to the URL identifier for that page. For example a file such as [en53003214/1903/05/01-ed-1-seg-1-ocr.txt](#) maps to the URL <http://chroniclingamerica.loc.gov/lccn/en53003214/1903-05-01-ed-1-seg-1->.

If you are interested in automated access to this data you may want to use the [Atom](#) and [JSON](#) versions of this table.

Please note: Some batches contain digitized newspaper pages that lack corresponding OCR text. For this reason, after decompressing the downloadable files, the number of .ocr.txt files may not match the page count available within the HTML Atom and JSON batch views.

Update (July 2019): In early June 2019, an error was identified with the [OCR Bulk Data download files](#) available from the [OCR report](#). These errors have been corrected but any files downloaded between February 5th and June 3rd 2019 should be replaced with the current versions. The source OCR data for all newspaper pages currently ingested in Chronicling America are also available at <https://chroniclingamerica.loc.gov/data/batches/>. Contact ndntech@loc.gov with any questions.

Filename **Batch** **Created** **Size** **SHA-1 Checksum**

vi_hanoverian_ver01.tarbz2	vi_hanoverian_ver01	2021-04-08T12:08:29-04:00	821.3 MB	06222c628993688bd9502a90f2af6809104160
sdbi_honeydew_ver01.tarbz2	sdbi_honeydew_ver01	2021-04-08T11:25:03-04:00	730.0 MB	409cb98680clb6a3441:07b8be599f72ce1c1778
njr_harris_ver01.tarbz2	njr_harris_ver01	2021-04-08T11:14:28-04:00	763.1 MB	ace7690a46973cd9e982e348aa799415ccb364bd
mimtptc_jerome_ver02.tarbz2	mimtptc_jerome_ver02	2021-04-08T11:04:41-04:00	296.6 MB	31e5a42c9d44cb0f2a133446d0b6037f33b9af
iahi_harivama_ver01.tarbz2	iahi_harivama_ver01	2021-04-08T09:46:04-04:00	860.1 MB	6b72a678b5ee9a9c27ab77678557471c2b60ab0
dic_unite_ver01.tarbz2	dic_unite_ver01	2021-04-08T09:30:30-04:00	802.2 MB	f16c4769e9bc7562e1e6ab8efc005934644bec4
dic_tartuffe_ver01.tarbz2	dic_tartuffe_ver01	2021-04-08T09:23:32-04:00	798.1 MB	030f168947c8daa2b280c7c942824ac02a684763
gu_oraco_ver01.tarbz2	gu_oraco_ver01	2021-04-07T18:57:47-04:00	1.3 GB	a64b3efde29373a3b03534fd67808698d2c92c
tvdn_gomez_ver01.tarbz2	tvdn_gomez_ver01	2021-04-09:04	906.0 MB	e6fcfa485d641ae35135c269a1da31e2c005a5ef

- Korisnici koji žele raditi s velikim brojem digitalnih datoteka i preuzeti ih na veliko, a ne pojedinačno, LC je osigurala obrađene podatke za preuzimanje naveliko npr. za:

- MARC Distribution Services osiguravajući sve LC zapise kataloga

[MARC Distribution Services \(data set\): Cataloging Distribution Service, Library of Congress \(loc.gov\)](#)

- National Digital Newspaper Program osigurava čisti OCR tekst novinskih stranica iz Chronicling America projekta

[OCR Data » Chronicling America » Library of Congress \(loc.gov\)](#)

- Jedan od ciljeva NDI-a je promicanje standarda i prakse oko pružanja pristupa zbirkama kao podacima tako da se puna vrijednost može računalno iskoristiti kako bi donijela još više svijest o znanju i kreativnosti sadržanoj u svjetskoj knjižnici.

PRISTUP DIGITALBIM ZBIRKAMA KAO PODACIMA

bulk download

primary digital resources **OPENN** available to everyone

Introduction - Technical Help - Repositories

OPenn contains complete sets of high-resolution archival images of cultural heritage material from the collections of its contributing institutions, along with machine-readable descriptive and technical metadata. All materials on OPenn are in the public domain or released under Creative Commons licenses as Free Cultural Works. Please see specific repository pages and documents for applicable license terms.

Name	Last modified	Size
CuratedCollections.html	2019-03-24 18:22	3.7K
Data/	2018-03-26 18:15	-
ReadMe.html	2018-03-24 18:22	13K
Repositories.html	2018-03-24 18:22	25K
TechnicalReadMe.html	2018-03-24 18:22	87K
collections.txt	2017-09-21 09:19	2.0K
logs/	2018-03-26 03:10	-
robots.txt	2015-09-10 14:54	24

tooled API



raw API

This repository Search

harvardartmuseums / api-docs

Code Issues Pull requests Projects Wiki Insights

Watch 7

API documentation for <https://www.harvardartmuseums.org>

153 commits 1 branch 0 releases

Branch: master New pull request Create new file Upload files Find

jeffsteward add initial documentation on errors

README.md add initial documentation on errors

activity.md change HAM api URLs from http to https

analysis.md change HAM api URLs from http to https

annotation.md add initial documentation on annotations

century.md change HAM api URLs from http to https

classification.md change HAM api URLs from http to https

color.md change HAM api URLs from http to https

culture.md change HAM api URLs from http to https

exhibition.md change HAM api URLs from http to https

experimental.md change HAM api URLs from http to https

gallery.md change HAM api URLs from http to https

aggregators' APIs



API - Chronicling America

- Search zapisi novinskih naslova i sadržaj digitaliziranih stranica se pretražuju upotrebom OpenSearch.
- Auto Suggest prilikom integracije u vlastitu aplikaciju bibliografski zapisi novinskih naslova pronalaze se upotrebom OpenSearch AutoSuggest API
- Link konstruiranje veza do određenih naslova novina, bilo kojeg dostupnog broja i njihovih izdanja, pa čak i do određenih stranica. Te se poveznice mogu lako označiti i dijeliti na drugim web mjestima.
- JSON dodatno upotrebi JSON-a u OpenSearch rezultatima, dostupan je i JSON prikaz za razne izvore u Chronicling America.
- Linked Data omogućuje povezivanje informacija u novinama izravno na povezane podatke na webu
- Bulk Data podaci su dostupni kao veliki paketi za istraživanja i vanjske usluge
- CORS and JSONP omogućuje integraciju novina u korisničke JavaScript aplikacije.

LIBRARY Digitization Services

ASK A LIBRARIAN DIGITAL COLLECTIONS LIBRARY CATALOGS Search Search Loc.gov

The Library of Congress > Chronicling America

Humanities CHRONICLING AMERICA Historic American Newspapers Search America's historic newspaper pages from 1789-1963 or use the U.S. Newspaper Directory to find information about American newspapers published between 1690-present. Chronicling America is produced jointly by the National Endowment for the Humanities and the Library of Congress. Learn More

Search Pages Advanced Search All Digitized Newspapers 1789-1963 US Newspaper Directory, 1690-Present

All states 1 from 1789 1 to 1963 1 + other site or more search terms

Pages Available: 15,006,843

100 Years Ago Today: 5/21/1919 (116 issues)

Print Subscribe Share/Save Give Feedback

About Chronicling America About the Site and API Recommended Topics Help

More Resources National Digital Newspaper Program NDNP Award Recipients Newspaper and Current Periodicals Reading Room Ask LC Newspaper & Current Periodicals Librarian Historic Newspapers on Flickr (part of the LC Flickr Commons photostream) Beyond Words (Help find pictures in historic newspapers!)

A screenshot of the Chronicling America website. The page displays a grid of historical newspaper pages from various years. One page from the Great Falls Daily Tribune in 1880 is shown in detail, featuring a headline about Wilson Wains Beer and Wins Back. Another page from the Belding Banner-News in 1860 is also visible. The interface includes search and filter options at the top.

DIGITALIZIRANE ZBIRKE

Open Data

<https://data.bnl.lu/data/historical-newspapers/>

Bibliothèque nationale du Luxembourg
Open Data

Historical Newspapers



Open Standards

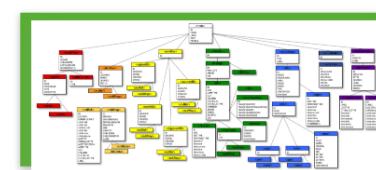


High Quality Data



Clear Copyright

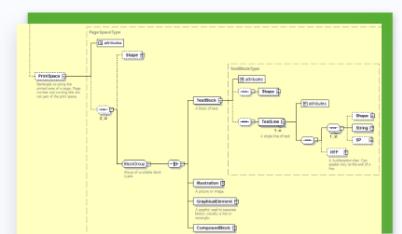
The screenshot shows a historical newspaper page from the Luxemburger Wort, dated January 23, 1877. The page features several columns of text in German, with some illustrations and a small sidebar at the top.



What is METS?

METS (Metadata Encoding and Transmission Standard) is a standard that allows the exchange of digitized documents between heritage institutions. It has been developed following the initiative of the Digital Library Federation (DLF) and is an implementation of the OAIS reference model (Model for an Open Archival Information System). Currently, the library of Congress in the United States of America is responsible for the maintenance of the METS schema. METS is an XML schema for the creation of digital objects. A digital object can be simple or complex, can consist of one or more digital files, which can be in different formats and describe detailed internal structure.

[Visit METS on the Library of Congress website](#) [Get METS Primer \(PDF\)](#)



What is ALTO?

ALTO (Analyzed Layout and Text Object) is an XML standard created as a result of the European project METAB and is designed to represent a physical document in terms of page layout, word positions and much more. It is used to store information about the content and layout of physical documents. In particular, it is especially well suited to represent OCR results. The Bnl uses ALTO together with METS. Each digitised page is represented by one ALTO file. ALTO files are responsible for the contents of individual pages and METS is responsible for the metadata, structural information and links between external files.

[Visit ALTO on the Library of Congress website](#)

SKUPOVI PODATAKA – OTVORENI PODACI

- Olakšavanje pristupa zbirkama za otvorenu znanost
- Optimizacija postojeće knjižnične ICT infrastrukture za stimulaciju održivih FAIR (Findable - Accessible - Interoperable - Reusable) pristupa na razini podataka za istraživanja digitalne humanistike

<https://data.bnllu/data/historical-newspapers>

Download a Dataset & Start Exploring!

Multiple datasets are available for download. Each one is of different size and contains different newspapers. All the digitised material can also be found on our search platform a-z.lu (Make sure to filter by "eluxemburgensis"). All datasets contain XML (METS + ALTO), PDF, original TIFF and PNG files for every newspaper issue.

STARTER PACK

250MB
of digitised newspapers

- ✓ 5 days of news
- ✓ 5 newspaper issues
- ✓ 22 pages
- ✓ DWäschfra (1868)
- ✓ Public Domain, CC0 (See copyright notice)
- ✓ Best for getting started & developing

[DOWNLOAD \(ZIP\)](#)

DEV PACK

3GB
of digitised newspapers

- ✓ 1 month of news
- ✓ 26 newspaper issues
- ✓ 112 pages
- ✓ Luxemburger Wort (1877)
- ✓ Public Domain, CC0 (See copyright notice)
- ✓ Best for getting started with Big Data

[DOWNLOAD \(ZIP\)](#)

SAMPLE PACK

1GB
of digitised newspapers

- ✓ 11 different newspaper titles
- ✓ 1 issue per newspaper
- ✓ News between 1845 and 1877
- ✓ Public Domain, CC0 (See copyright notice)
- ✓ Best for testing different newspapers and metadata

[DOWNLOAD \(ZIP\)](#)

Large Datasets

For advanced users, the BnL provides larger datasets. Those datasets are meant for data scientists and researchers, especially in the field of digital humanities where large quantity of data can train machine learning algorithms or neural networks. Note that you can combine all datasets to form an even larger set.



32GB
of digitised newspapers

- ✓ 1 year of news
- ✓ 304 newspaper issues
- ✓ 1220 pages
- ✓ L'indépendance Luxembourgeoise (1877)
- ✓ Public Domain, CC0 (See copyright notice)
- ✓ Best for getting started with machine learning

[DOWNLOAD \(ZIP\)](#)

257GB
of digitised newspapers

- ✓ 10 years of news
- ✓ 2712 newspaper issues
- ✓ 10880 pages
- ✓ L'Union (1860-1869)
- ✓ Public Domain, CC0 (See copyright notice)
- ✓ Best for machine learning and deep neural networks

[DOWNLOAD \(ZIP\)](#)

Processed Datasets

Working with the raw data can be tedious. For that reason, the BnL processed all newspapers and monographs that are in the public domain and extracted the full text and associated meta data of every single article, section, advertisement... The result is a large number of small, easy to use XML files formatted using Dublin Core.

The open source tool of the BnL has been used to create the export. [Documentation of the format](#)

TEXT ANALYSIS PACK

2GB
of processed newspapers data

- ✓ 38 years of news (1841-1878)
- ✓ 23663 processed newspaper issues
- ✓ 97285 processed pages
- ✓ 51050 extracted articles
- ✓ Public Domain, CC0 (See copyright notice)
- ✓ Best for getting started with text analysis

[DOWNLOAD \(ZIP\)](#)

MONOGRAPH TEXT PACK

125MB
of processed monographs data

- ✓ 228 years period (1690-1918)
- ✓ 504 processed monographs
- ✓ 51709 processed pages
- ✓ 33477 extracted chapters
- ✓ Public Domain, CC0 (See copyright notice)
- ✓ Best for getting started with text analysis

[DOWNLOAD \(ZIP\)](#)

Collections as Data – LC: Newspaper Navigator Dataset

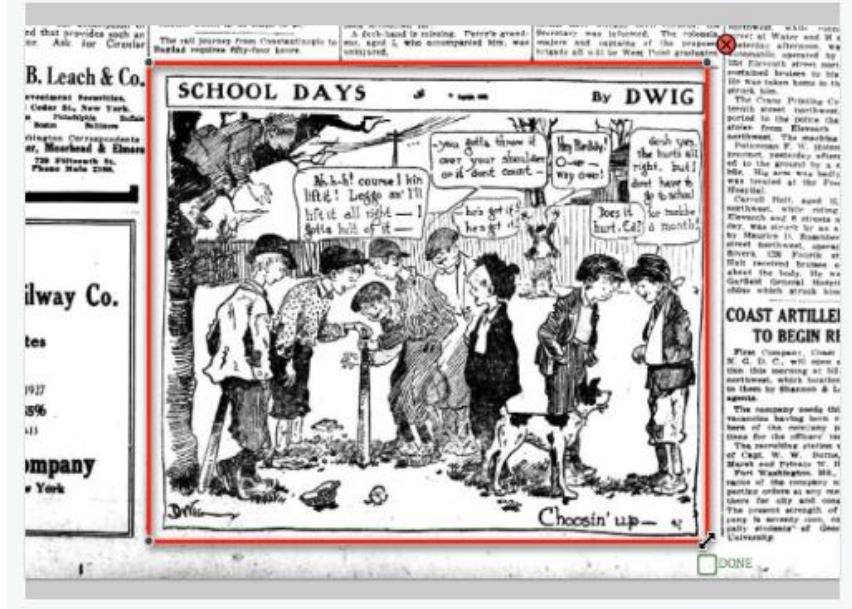


Visual content type	ZIP w/ JPEGs	JSON file w/ metadata	CSV file w/ metadata	Visual content type	JSON file w/ metadata	JSON file w/ embeddings
Photos	1905_photos.zip	1905_photos.json	1905_photos.csv	Photos	1905_photos_sample.json	1905_photos_sample_embeddings.json
Illustrations	1905_illustrations.zip	1905_illustrations.json	1905_illustrations.csv	Illustrations	1905_illustrations_sample.json	1905_illustrations_sample_embeddings.json
Maps	1905_maps.zip	1905_maps.json	1905_maps.csv	Maps	1905_maps_sample.json	1905_maps_sample_embeddings.json
Comics	1905_comics.zip	1905_comics.json	1905_comics.csv	Comics	1905_comics_sample.json	1905_comics_sample_embeddings.json
Editorial Cartoons	1905_cartoons.zip	1905_cartoons.json	1905_cartoons.csv	Editorial Cartoons	1905_cartoons_sample.json	1905_cartoons_sample_embeddings.json
Headlines	N/A	1905_headlines.json	1905_headlines.csv	Headlines	1905_headlines_sample.json	N/A
Advertisements	N/A	1905_ads.json	1905_ads.csv	Advertisements	1905_ads_sample.json	1905_ads_sample_embeddings.json

LC - Beyond Words: MARK – TRANSCRIBE – VERIFY

Task 1: Mark

Here you'll be identifying various types of pictures (illustrations, photographs, and cartoons) in the newspapers. You won't be typing, just drawing boxes. You may see existing marks left by other users. Just pick up where they left off. Do not identify pictures within an advertisement.



Task 2: Transcribe

You'll enter the text inside the fields marked by your fellow volunteers. Each transcribe window has 3 fields: Caption, Artist, and Category.

The Caption field is the place to transcribe the title, caption, and cutline of the picture. Some caption fields may be prepopulated with OCR (Optical Character Recognition) text. From time to time you may find OCR text inaccurate and need to be corrected. You can also delete the OCR text and transcribe the caption. Enter the text exactly as you see it. Include capitalization and punctuation, but remove hyphenation that breaks words at the end of the line. Use new lines to separate different parts of captions and descriptions.

The Artist field is where you will enter the illustrator, photographer, cartoonist, or artist name – if present on the picture. Please enter as written. You'll use the Category field to tell us what kind of picture is being transcribed: Editorial Cartoon, Comics/Cartoon, Illustration, Photograph, or Map.

Sometimes you may find the text too small to read. You can zoom using the controls in the upper left corner of the screen; you can also zoom with keyboard commands: (**Control** + **Alt** + **+**) and pan using: (**Control** + **Alt** + **1** / **2** / **3** / **4**) or (**Control** + **Alt** + **W** / **E** / **N** / **S**). It may also be helpful to view the full-resolution newspaper page via the "View the original page" link on the right panel. If you still find it hard to decipher, don't worry. Together we'll transcribe these pages several times to ensure accuracy, so give it your best shot and move on.

Original prompt: Enter caption (or article title if caption is absent), creator and category.

Caption: SCHOOL DAYS

Creator: DWIG

Category: Illustration

Task 3: Verify

Every newspaper page will be transcribed several times by multiple volunteers, and our software looks for agreement among them. If a field has been transcribed consistently and independently multiple times, it has reached "consensus" and is added to the final data sets.

However, conflicting transcriptions are queued up in the Verify task. Here's where we need your help: you can select the most accurate transcription. If none of the options appear correct, you can also add your own.

Choose the best transcription or enter your own!

Original prompt: Enter caption (or article title if caption is absent), creator and category.

caption: SCHOOL DAYS

creator: DWIG

category: Illustration

caption: SCHOOL DA HIRL cts (\$ until 1 of 2) DWIG hnt, h^2 a month y^6

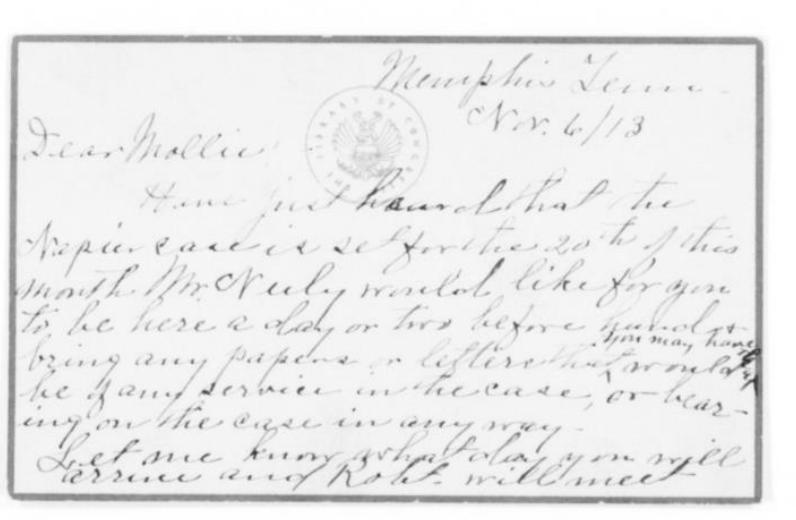
creator: DWIG

category: Illustration

None of these look your own.

LC: BY THE PEOPLE

Page 2 Go < > Fullscreen [View on www.loc.gov](#)



✓ Completed
This transcription is finished! You can read and add tags.

Memphis Tenn
Nov. 6/13

Dear Mollie

Have just heard that the Napier case is set for the 20th of this month. Mr. Neely would like for you to be here a day or two before hand & bring any papers or letters that you may have that would be of any service in the case, or bearing on the case in any way.

Let me know what day you will arrive and Robt. will meet

Tags

Want to tag this page? Register or [login](#) to add tags.

Napier case ×

Need help?

Quick Tips

Questions?

THE REAL FACE OF WHITE AUSTRALIA

the real face of white australia

home • about



Form No. 21. COMMONWEALTH OF AUSTRALIA. No. 06/213
Immigration Restriction Acts 1901-1905 and Regulations.

CERTIFICATE EXEMPTING FROM DICTATION TEST.

I, *Nicholas L. Lockett*, the Collector of Customs for the State of *New South Wales* in the said Commonwealth, hereby certify that *John Jones*, hereinafter described, who is leaving the Commonwealth temporarily, will be excepted from the provisions of paragraph (a) of Section 3 of the Act if he returns to the Commonwealth within a period of *three years* from this date.

Date *5 June 06* Collector of Customs.

DESCRIPTION

Nationality *Chinese* Birthplace *Canton*
Age *39 years* Complexion *Dark*
Height *5 ft 5 3/4 in in Boots* Hair *Dark*
Build *Medium* Eyes *Brown*
Particular marks *Slight scar on right cheek bone.*

(For impression of hand see back of this document.)

PHOTOGRAPHS.

Profile: —

By Authority: Rom. S. Evans, Government Printer, Melbourne.

AUSTRALIA H. M. S. N. S. W. SYDNEY Date of departure *June 06* Destination *China*
Ship *Empire Empress*

Date of return *16^a August 1907* Ship *Empire Empress*
Port *Sydney*

Customs Officer *W. J. D. Morris*

ZAKLJUČAK

- Korisnici od AKM zajednice očekuju pristup digitalnim zbirkama kao podacima koji su im potrebni za njihov istraživački, obrazovni ili kreativni rad
- Poticati razvoj zbirki kao podataka u AKM zajednici Hrvatske
- Poticati razvoja laboratorija u AKM zajednici – mesta suradnje s mogućnošću umrežavanja unutar AKM zajednice
- Posebnu pažnju posvetiti ostvarivanju partnerstva sa sveučilištima i znanstvenim institucijama, znanstvenicima, nastavnicima, studentima i umjetnicima
- Stvaranje interdisciplinarnih timova koji će raditi i koristiti zbirke i podatke
- Koordinirano uvesti tečajeve za AKM zajednicu o digitalnim metodama i alatima koji će pomoći uvesti koncept korištenja knjižničnih zbirki kao podataka kroz razvoj potrebnih vještina, znanja i resursa