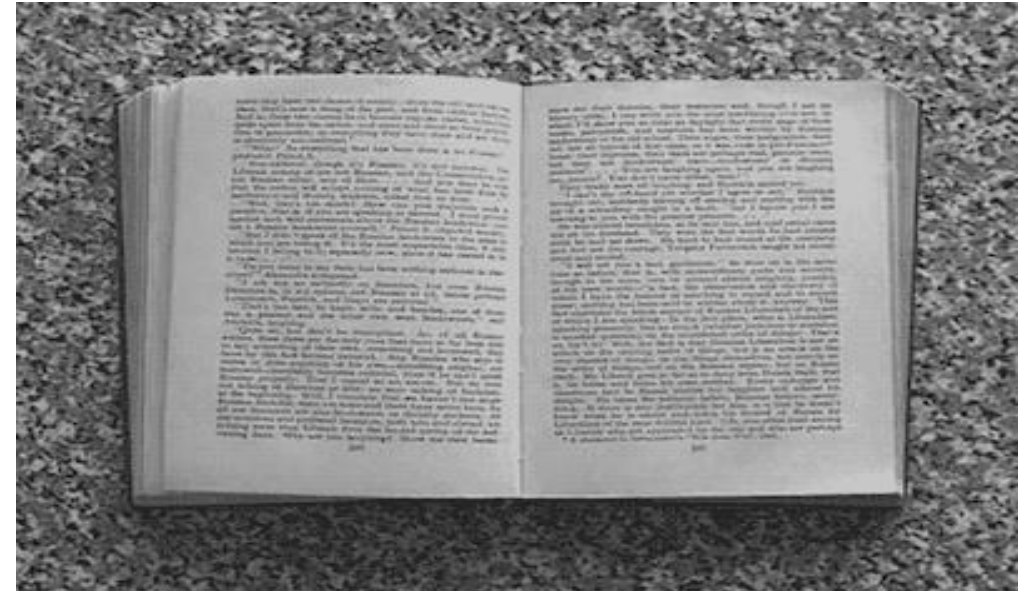
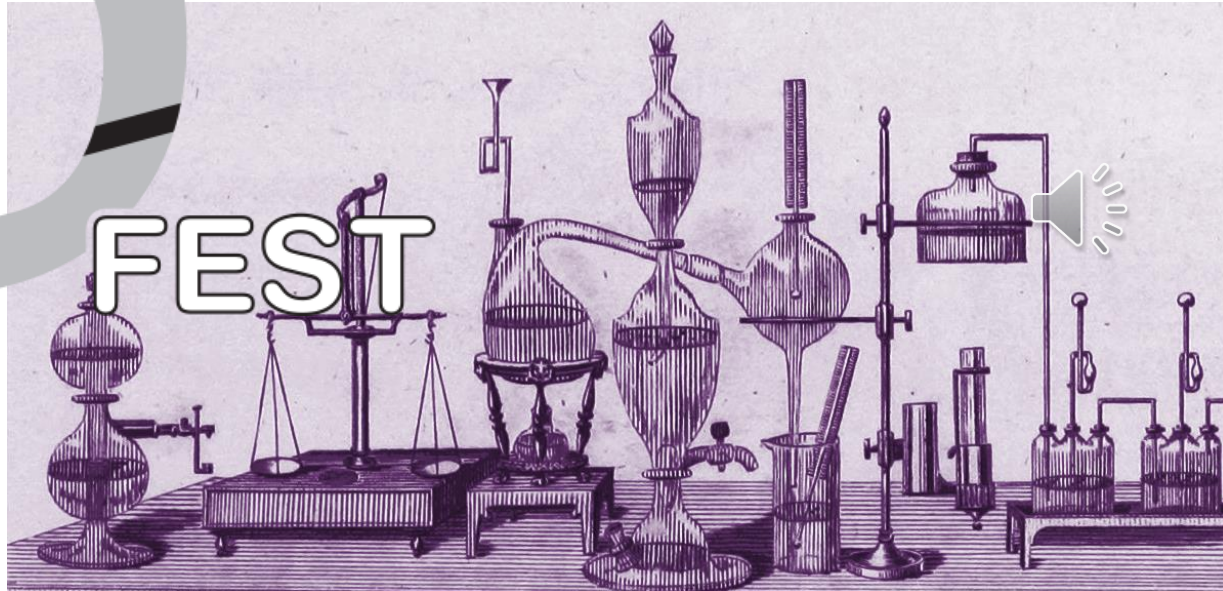


Digitalno objedinjavanje kroz razvoj „zbirki kao podataka“ i baštinskih suradničkih inovacijskih laboratorija



FEST



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 orijentiran 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 tijekom rada na korisnički pristupačan način**
- prilagoditi digitalizirane zbirke za korištenje digitalnih humanističkih metode 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 pomoć 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 access and use of digital collections, 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)

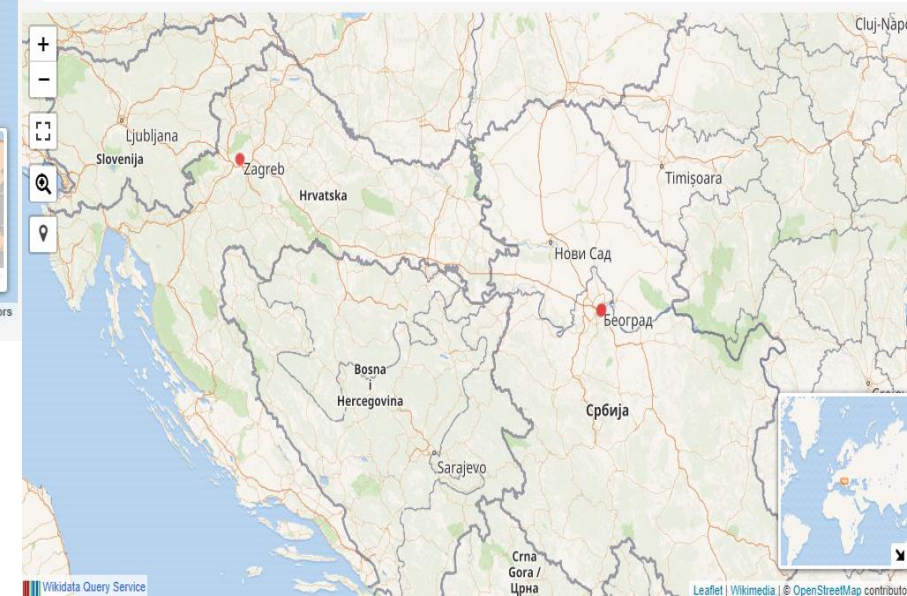
Dutch KB Lab – osnovan 2014,

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>

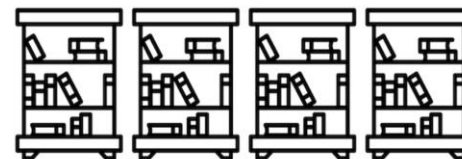
ZBIRKE KAO PODACI

- Bez podataka nema laboratorija, stoga digitalne zbirke trebaju biti dostupne u formi podataka - temeljnoj aktivnosti laboratorija
- Transformacija zbirke 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 zbirka 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 zbirke kulturne baštine. Zbirke kao podaci postavljaju pitanje što bi moglo značiti tretirati digitalizirane i izvorne digitalne zbirke kao podatke, a ne kao jednostavne surogate 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 zbirka i velikim paketima otvorenih podataka i metapodataka povezanih s digitaliziranim i izvornom digitalnim zbirka 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 ensuring the future of collections that 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



212-11 357.455
 BRLIĆ-MAŠURANIĆ, IVANA
 Ivana Brlić-Mašuranić. Priče iz davnine.
 (Ilustr. Albert Kinert.) 2. izd. (u
 lioteci).
 Zagreb, Mladost, 1961. 8* 1
 sa sl.
 (Biblioteka Vjeeverica.
 Sv. 36.)

165.6+141.4+141.131>(082)
 583.634

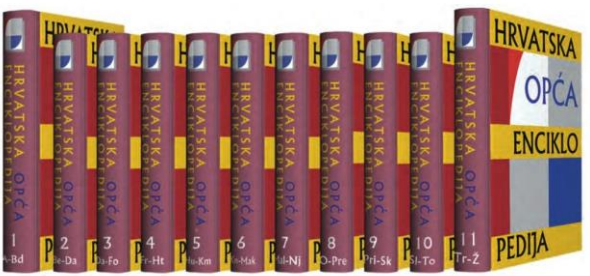
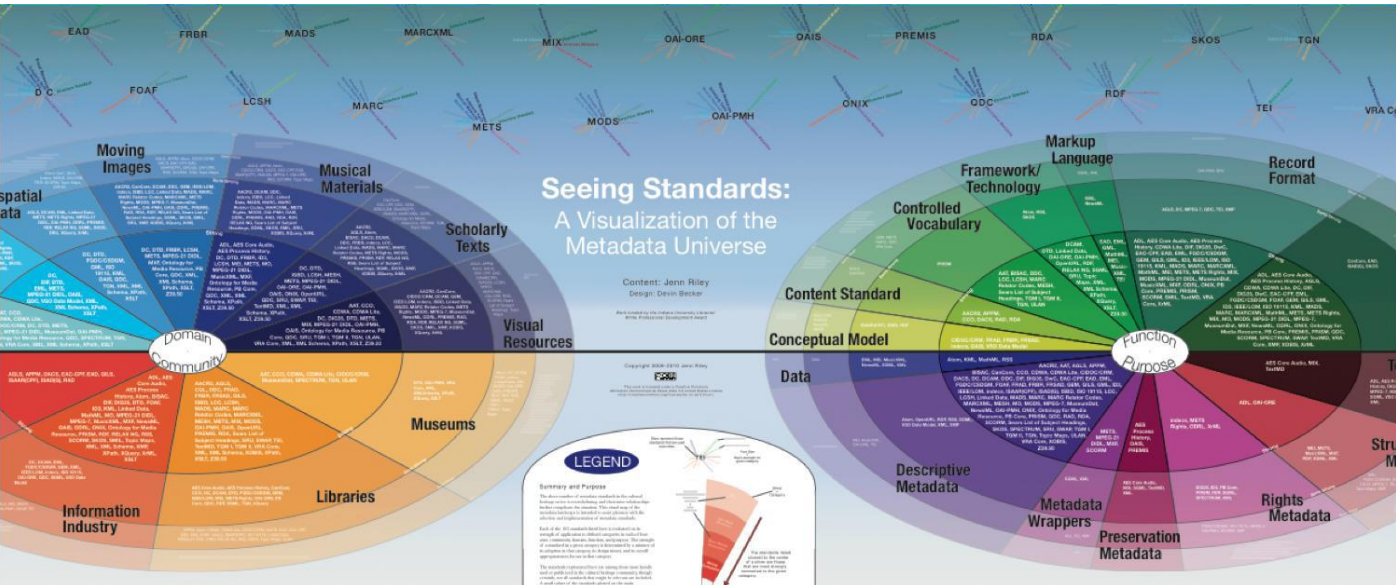
OD Galileja do Platona. - Ljubljana :
 RK ZSMS, <1986> (<Ljubljana> : DDU
 Univerzum). - 183 str. : graf. prikazi ;
 20 cm. - (Zbirka Analacta)

Om. nasl. - Izredna številka revije
 Problemi. - Sadržaj: Elementa Galilejske
 epistemologije / Miran Božović. Bog
 je nezaveden / Francois Regnault. Plato-
 nistična paralaksa / Igor Zagar. Franco-
 ska zveza / Marce Stefančič.

11/15/901407 4



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 ... zbirke kao podataka. (Thomas Padilla, 2016)



PRISTUP DIGITALBIM ZBIRKAMA KAO PODACIMA

bulk download



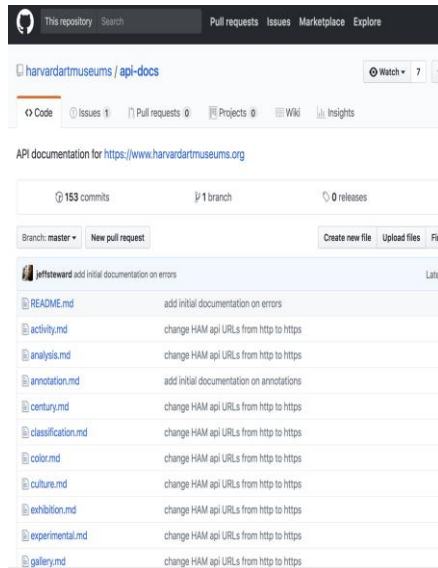
OPENN contains complete sets of high-resolution archival images of cultural heritage material from the collections of its contributing institutions, along with machine-readable descriptive and technical metadata. All materials on OPENN are in the public domain or released under Creative Commons licenses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

Name	Last modified	Size
CuratedCollections.html	2018-03-24 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

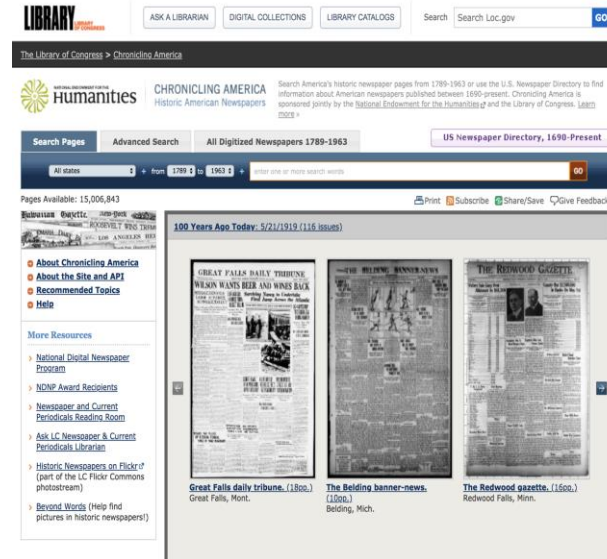


aggregators' APIs



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.




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

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 "eluxemburgensia"). 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)

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

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
- 510505 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.



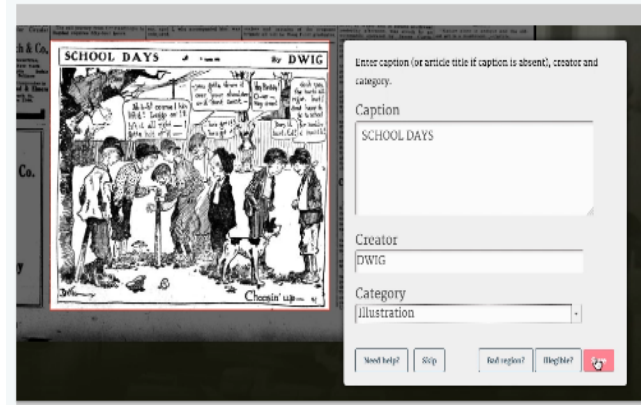
Task 2: Transcribe

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 outline 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.

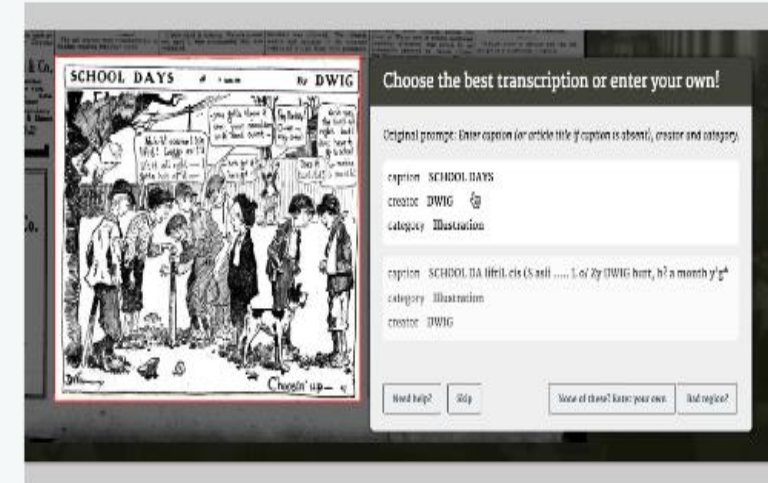
Sometimes you may find the text too small to read. You can zoom using the controls in the upper left corner of the screen; you can also zoom with keyboard commands: **Control** **Alt** **+** **[** and pan using: **Control** **Alt** **+** **←** or **Control** **Alt** **+** **→**. It may also be helpful 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.



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.



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 x

Need help? Quick Tips Questions?



[By the People mss425490015-2 \(Mary Church Terrell: Advocate for African Americans and Women: Family letters\) \(loc.gov\)](#)

THE REAL FACE OF WHITE AUSTRALIA

the real face of white australia

home • about



the real face of white australia

The White Australia Policy was about people – people whose lives were monitored and restricted because of the colour of their skin. This experimental browser enables you to explore the records of the White Australia Policy through the faces of those people.

These portraits were extracted from a range of government documents using a [face detection script](#). We've tried to weed out the mistakes, but you may still notice a few oddities. Many portraits are duplicated, as multiple copies of the forms were often kept.

The records are held by the [National Archives of Australia](#). Currently the browser only shows images from Series [ST84/1](#). Other series will be added over time. You can read more about the records at [The Tiger's Mouth](#).

If you click on a portrait you can view the document it came from. You'll also be able to follow a link to explore the context of the document in the National Archives of Australia's [RecordSearch](#) database.

The portraits are presented in random order. You can reverse the order simply by adding `?order=reverse` to the url. You can also browse file by file by adding `?order=file`.

This experimental browser has been developed as part of the [Invisible Australians](#) project.

Invisible Australians

Form No. 2d. **COMMONWEALTH OF AUSTRALIA.** No. *06/213*
Immigration Restriction Acts 1901-1905 and Regulations.

CERTIFICATE EXEMPTING FROM DICTATION TEST.

I, *Nicholas Le Cocq* the Collector of Customs for the State of *New South Wales* in the said Commonwealth, hereby certify that *Ah Jow* hereinafter described, who is leaving the Commonwealth temporarily, will be excepted from the provisions of paragraph (d) 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* *Yorkie* Collector of Customs.

DESCRIPTION	
Nationality <i>Chinese</i>	Birthplace <i>Canton</i>
Age <i>39 years</i>	Complexion <i>Dark</i>
Height <i>5ft 5 3/4 in in Boots</i>	Hair <i>Dark</i>
Build <i>Medium</i>	Eyes <i>Brown</i>
Particular marks <i>Slight scar on right cheek bone.</i> (For impression of hand see back of this document.)	

PHOTOGRAPHS. Profile:—

By Authority: *Roger S. Evans*, Government Printer, Melbourne.

Date of departure *June 06* Destination *China*
Ship *Comore*
Date of return *16th August 1907* Ship *Empire*
Port *Sydney* *W. J. Amokor* Customs Officer

ZAKLJUČAK

- Korisnici od AKM zajednice očekuju pristup digitalnim zbirkama kao podacima koji su im potrebni za njihov istraživački, obrazovni ili kreativni rad
- Poticati razvoj zbirke kao podataka u AKM zajednici Hrvatske
- Poticati razvoja laboratorija u AKM zajednici – mjesta 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 zbirke kao podataka kroz razvoj potrebnih vještina, znanja i resursa