Practical Digital Curation Skills for Archivists in the 21st Century

Myeong **LEE**
Mary **KENDIG**
Richard **MARCIANO**
Greg **JANSEN**

U. Maryland iSchool DCIC

MARAC FALL 2016

ANNAPOLIS, MD | NOVEMBER 3-5, 2016

UNIVERSITY OF MARYLAND

COLLEGE OF INFORMATION STUDIES

dcic digital curation innovation center

Agenda

- Introduction to DCIC
- Defining Digital Curation
- Case Studies
    1. Human Face of Big Data [Community Displacement]
    2. Mapping Inequality [Racial Zoning]
    3. St. Louis Voyage [Refugee Narratives]
    4. WWII Japanese American Camps [Citizen Internment]
- Computational Archival Science Themes
- Practical Skills for 21st Century Archivists
- Conclusion / Questions

ARC: Archives Research & Collaboration Lab

Director: Ricky **Punzalan**

ARC studies and develops innovative approaches, systems, strategies, and tools to foster sustainable futures for archives, preservation, and digital curation.

http://archivescollaboratory.umd.edu/

SALT: Sustainable Archives & Leveraging Technologies

Director: Richard **Marciano**

SALT is an interdisciplinary lab, which focuses on the long-term preservation of digital cultural and research assets at scale. SALT is an acronym for SustainABiLiTy and uses as its logo the two thousand year-old ancient Chinese pictograph for salt ("yan") which is a metaphor for the integration of policy, governance, infrastructure, and content.

http://salt.umd.edu

curatelab

Hornbake South 4110

Digital lab for group learning, collaborative design, and hands-on digital curation project development (23 seats, 3 interactive screens, 12 workstations with 12TB of storage).

digitizationlab

Hornbake South 4110D

Document scanning, image manipulation, and archival ingestion facility for group projects.

serverfarm

UMD Computer & Space Sci. Bldg

On-campus virtual machine farm for research data processing, storage, and hosting (15TB storage, 2 Dell servers, VMWare-powered).

cloudlab

Amazon Cloud

Dashboard-enabled virtual computing lab in the cloud for creating Windows/Ubuntu instances using Amazon Web Services (AWS).

dataCave

UMD Cyberinfrastructure Center at the Rivertech Bldg

Building **DRAS-TIC**

Digital Repository At Scale That Invites Computation (To Improve Collections): a peta-scale archival storage and preservation repository (based on **DRAS-TIC** open-source software [NoSQL Cassandra database] and computational infrastructure (4 Dell nodes).

dcic

digital curation innovation center

http://dcic.umd.edu

Mission

Be a leader in the digital curation research and educational fields, and foster interdisciplinary collaborations using Big Records and archival analytics with public / industry / government partnerships.

Sponsor interdisciplinary projects that explore the integration of archival research data, user-contributed data, and technology to generate new forms of analysis and historical research engagements, particularly in the arenas of social justice, human rights, and cultural heritage

UNIVERSITY OF MARYLAND

COLLEGE OF INFORMATION STUDIES

Projects

Cyberinfrastructure for the curation & management of digital assets at scale:

"Brown Dog"

A CIC Big-10 $10.5M NSF/DIBBs-funded collaboration with U. of Illinois NCSA Supercomputing Center and industry partners (NetApp and Archive Analytics Solutions). This project aims to help accelerate the development of digital curation processes and services and create a data observatory to provide access to Big Records training sets and teach students practical digital curation skills.

"Curate Cloud"

A $300K IMLS-funded project that helped launch a new online professional education certificate for digital curation professionals, the Curation and Management of Digital Assets (CMDA). Curate Cloud is also developing an open-source research and educational platform, the VCL (Virtual Computing Lab), to remove barriers to access for curation tools and resources.

Digital Curation training:

Digital Curation Fellowships

The iSchool has several Fellowship opportunities for students in digital curation and archives. These include a collaboration with the **National Agricultural Library** (NAL); extensive project work with the **National Park Service** (NPS); and a scholarship established in honor of **Bruce Ambacher**, retired senior archivist and iSchool faculty member.

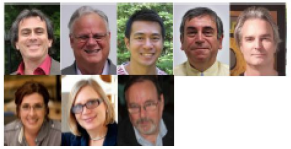
Interdisciplinary Research Teams

Gain new digital skills, conduct interdisciplinary research, explore professional development opportunities at the intersection of archives, big data, and analytics through a number of project themes: *Refugee Narratives, Community Displacement, Racial Zoning, Cyberinfrastructure for Digital Curation, Movement of People, Citizen Internment.*

People

Research Staff:

Richard **Marciano**	*Director & SALT Lab Director*
Michael **Kurtz**	*Associate Director*
Ricardo **Punzalan**	*Research Associate & ARC Lab Director*
Ken **Heger**	*Research Associate & DigitizationLab Director*
Greg **Jansen**	*Research Software Architect*
Maria **Esteva**	*Affiliate Professor*
Victoria **Lemieux**	*Affiliate Professor*
William **Underwood**	*Affiliate Professor*



Mary **Kendig** / Myeong **Lee** *Graduate Research Assistants*

Research Affiliates:

U. Maryland

Tammy **Clegg**, Nick **Diakopoulos**, Jesse **Johnston**, Trevor **Owens**, Jenny **Preece**, Katie **Shilton**

External

Bruce **Ambacher**, Natalie **Baur**, John **Burns**, Andrew **Lau**, Scott **Madry**

Postdoctoral Fellows:

Morgan **Daniels**, Kathryn **Gucer**, Adam **Kriesberg** *(Advisor: Punzalan)*

Students (Undergraduate, Master's [MLIS, MIM, HCIM], Doctoral):

Maddie **Allen**, Saba **Al-Dughaither**, Vinita **Atre**, Myuresh **Amdekar**, Richard **Bool**, Arpit **Chandra**, Shiyun **Chen**, John **Dela Cruz**, Anne **Dempsey**, Shaina **Destine**, Kelsey **Diemand**, Pal **Doshi**, Erin **Durham**, Will **Froliklong**, Alicia **Geller**, Karishma **Ghiya**, Janet **Glazier**, Rajesh **Gnanasekaran**, Rhett **Greenfield**, Allison **Gunn**, Ashley **Haddix**, Scott **Harkless**, Tonra **Hausmann**, Eric **Hung**, Hardik **Jhaveri**, Ruchira **Kapoor**, Amar **Kurane**, Yuting **Liao**, Zhenye **Ma**, Sheryl **Mathias**, Paridhi **Mathur**, Martin **Moreno**, Jennifer **Proctor**, Brian **Redford**, Darlene **Reyes**, Benjamin **Sagey**, Sohan **Shah**, Jay **Sheth**, Niraj **Shirame**, Edel **Spencer**, Akash **Udani**, Sydney **Valle**, Jennifer **Wachtel**, Melissa **Wertheimer**, Meaghan **Wilson**, Jiahui **Wu**, David **Zhang**, Xinyun **Zhang**

Doctoral Students:

Andrew **Casertano**, Will **Thomas**	*(Advisor: Marciano)*
Diane **Travis**	*(Advisors: Butler/Marciano)*
Edward **Summers**, Amy **Wickner**	*(Advisor: Punzalan)*

Projects

Justice, Human Rights, & Cultural Heritage:

Overseas Pension Project

A student- and professional society-driven project to collect information documenting payment of pensions to American veterans living overseas. The project creates datasets documenting migration patterns, the flow of money, health conditions, and family connections prior to World War I.

International Research Portal Project (IRP²)

This project will improve access to an important tool which identifies and locates looted art and other cultural assets found on the *International Research Portal for Records Related to Nazi-Era Cultural Property.*

Mapping the Voyage of the St. Louis

In 1939, 937 passengers (mostly Jews) fled Germany aboard the SS St. Louis ship, heading to Cuba, where they were turned away and forced to return to Europe where 254 were killed during the Holocaust. The project looks at mapping individual and collective stories through graph database techniques.

Japanese-American WWII Camps

Building on a UMD FIA Seed Grant, the project explores the integration of archival and user-contributed data using social networking graphs to link people, places, and events. Using WWII Camp data.

Mapping Inequality

A project with Johns Hopkins, Virginia Tech, and U. of Richmond where a national collection of New Deal *redlining* records is being crowdsourced (these unique records capture racial, ethnic, and economic conditions).

The Human Face of Big Data

A student-led project that will create access and collaborative opportunities around historically and socially- significant heterogeneous datasets rooted in urban renewal housing records for a number of cities.

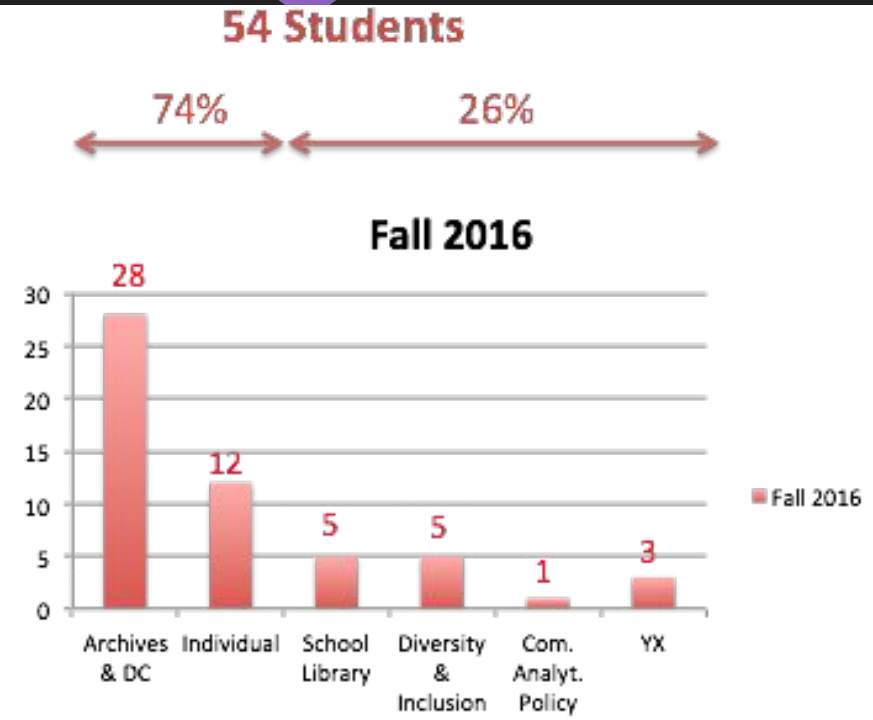
Digital Curation Innovation Center at UMD

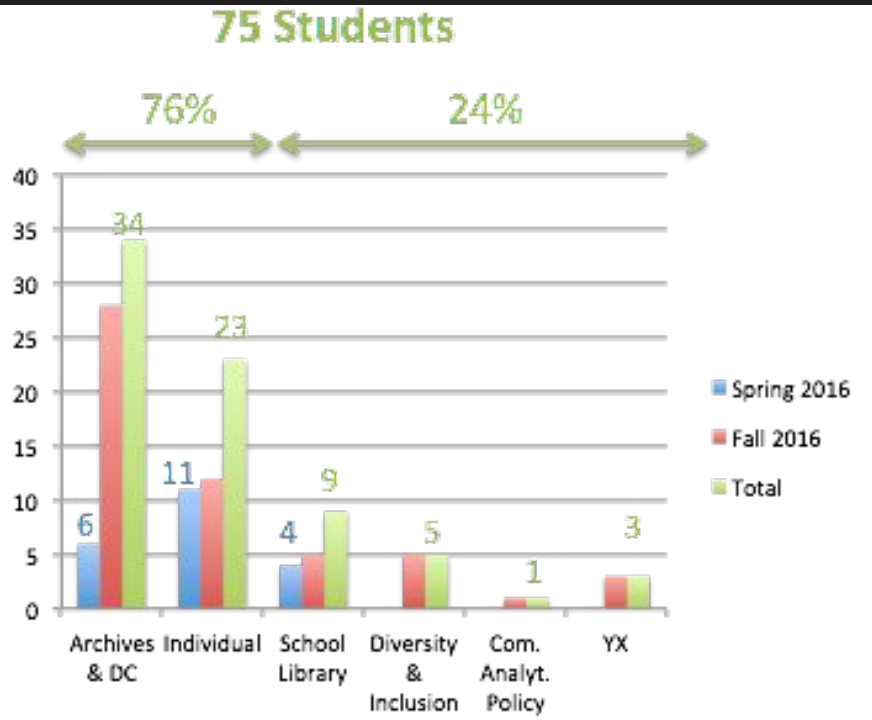
Mission:

- Be a leader in the digital curation research and educational fields, and foster interdisciplinary partnerships using **Big Records and archival analytics** through public / industry / government collaborations.

- Sponsor interdisciplinary projects that explore the integration of archival research data, user-contributed data, and technology to **generate new forms of analysis and historical research engagements.**

U. Maryland iSchool MLIS Enrollment Stats

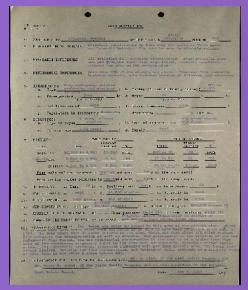


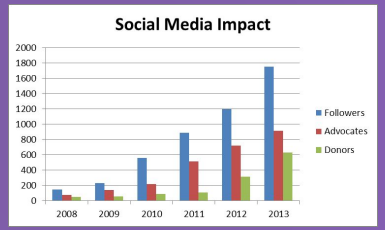


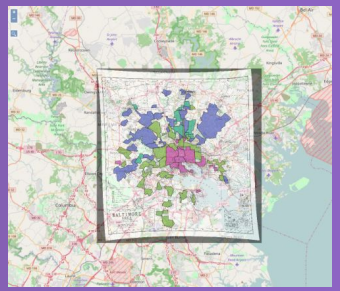
Digital Curation…

“The active and ongoing management and enhancement of digital assets for current and future use.” Digital curation entails more than secure storage and preservation of digital information because curation may add value to digital information and increase its utility.

[Preparing the Workforce for Digital Curation (2015) - NRC / BRDI Report]







Archival Documents **Digitization** **Datafication Data Modeling** **Data Visualization Digital Storytelling**

1. Human Face of Big Data

[Community Displacement]

UMD Student Team:

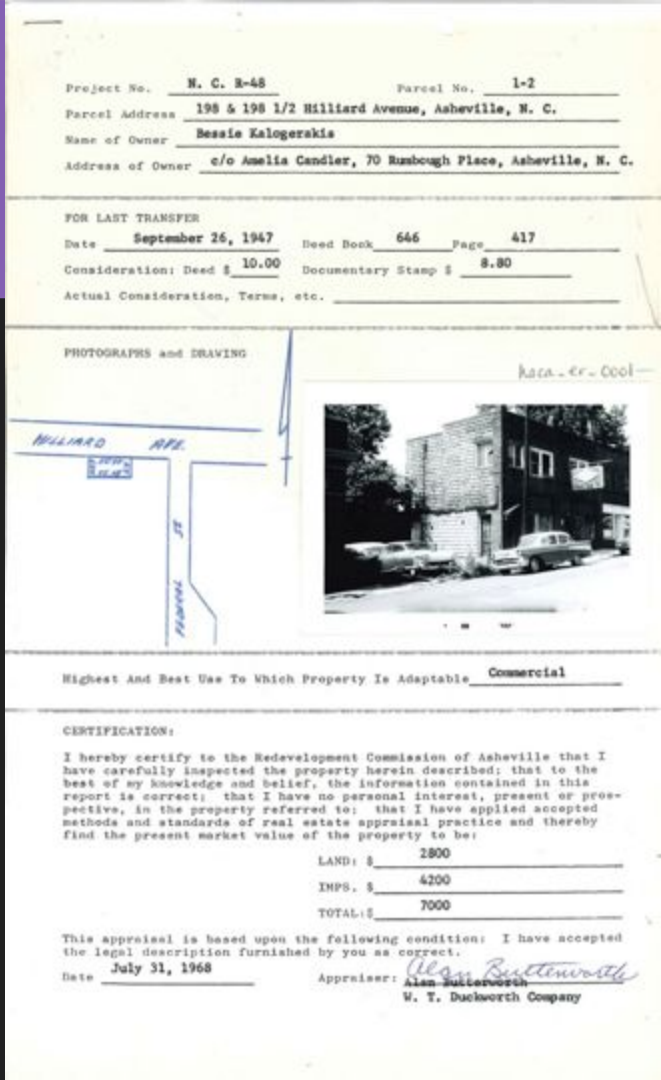
Myeong Lee
Shiyun Chen
David Zhang
Edel Spencer
Rajesh Gnanasekaran
Alicia Geller
Hardik Jhaveri

Urban Renewal Documents

How each property was acquired by Housing Authority?

How did the old neighborhood look like?

Providing a forest of how the urban renewal project was conducted in a city through building a system.



Project No. N. C. R-48 Parcel No. 1-2

Parcel Address 198 & 198 1/2 Hilliard Avenue, Asheville, N. C.

Name of Owner Bessie Kalogerakis

Address of Owner c/o Amelia Candler, 70 Rumbough Place, Asheville, N. C.

FOR LAST TRANSFER

Date September 26, 1947 Deed Book 646 Page 417

Consideration: Deed $ 10.00 Documentary Stamp $ 8.80

Actual Consideration, Terms, etc.

PHOTOGRAPHS and DRAWING

HILLIARD AVE.

Highest And Best Use To Which Property Is Adaptable Commercial

CERTIFICATION:

I hereby certify to the Redevelopment Commission of Asheville that I have carefully inspected the property herein described; that to the best of my knowledge and belief, the information contained in this report is correct; that I have no personal interest, present or prospective, in the property referred to; that I have applied accepted methods and standards of real estate appraisal practice and thereby find the present market value of the property to be:

LAND: $ 2800

IMPS. $ 4200

TOTAL: $ 7000

This appraisal is based upon the following condition: I have accepted the legal description furnished by you as correct.

Date July 31, 1968

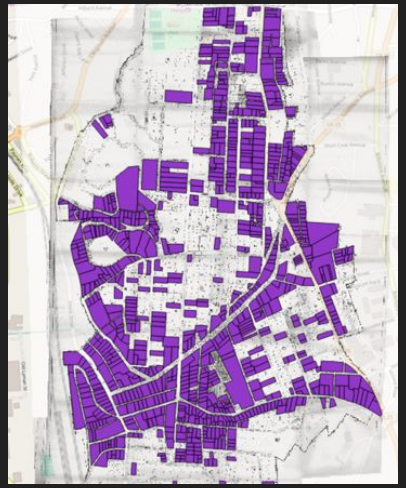
Appraiser: Alan Butterworth
W. T. Duckworth Company

Digitization Process of an Old Map

Paper Map

Scanning & Adjusting

Georeferencing



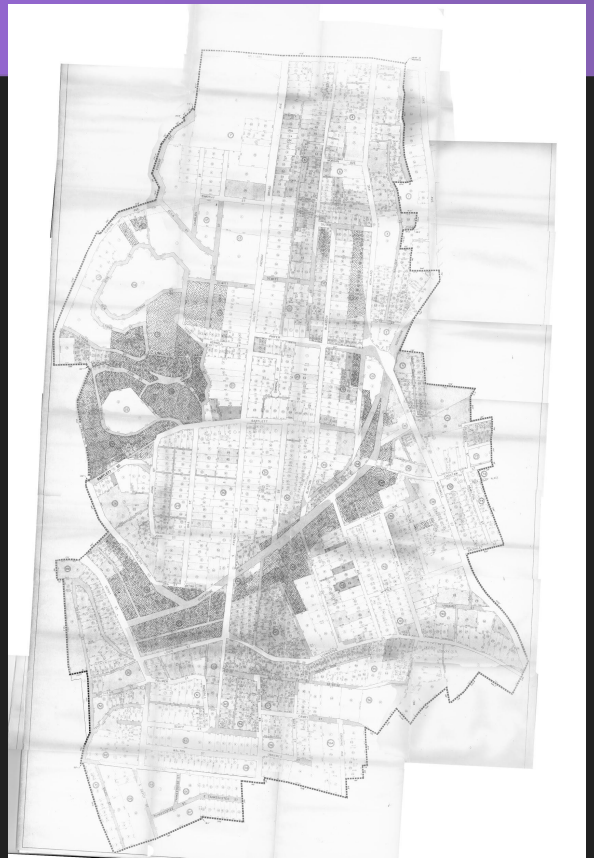
Geo-tracing

Scanning and Adjusting

Distortions Due to Tiled Scanning

Smart Object in Photoshop

Georeferencing



Using ArcGIS or QGIS

Aligning an Old Raster Map onto the Modern Online Map

Used another clean Map as a Guideline to Align the Scanned Map

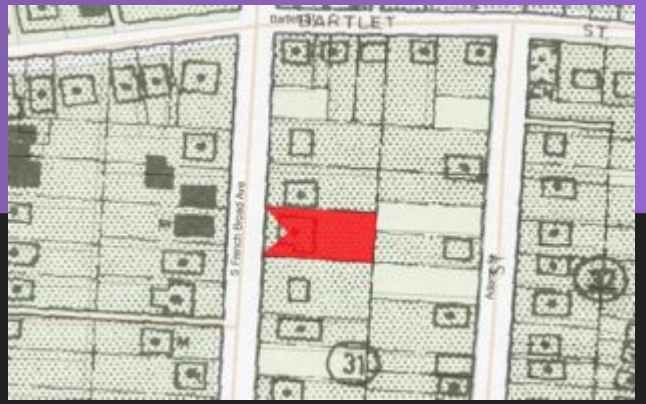
Geo-Tracing

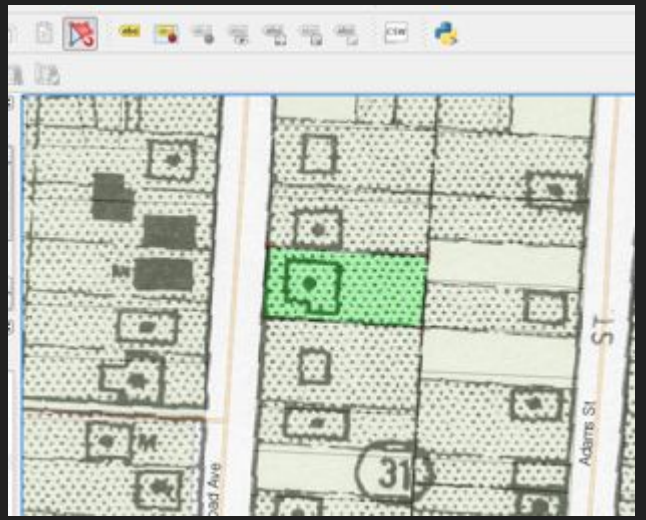
The process of creating **computer-detectable polygons** on top of the raster map.

Possible using **QGIS**, **ArcGIS**, or **Leaflet.draw**.

Stored in a **Shapefile**.

Shapefile can be exported to **GeoJSON**.





Iterative System Design and Prototyping

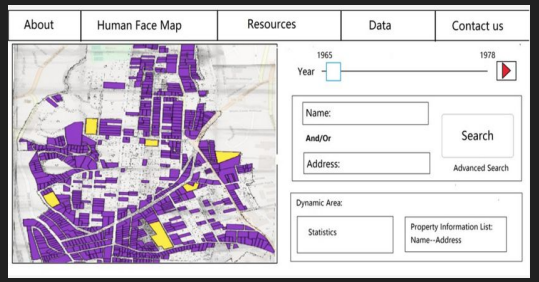
User Persona

User Scenario

Data Modeling

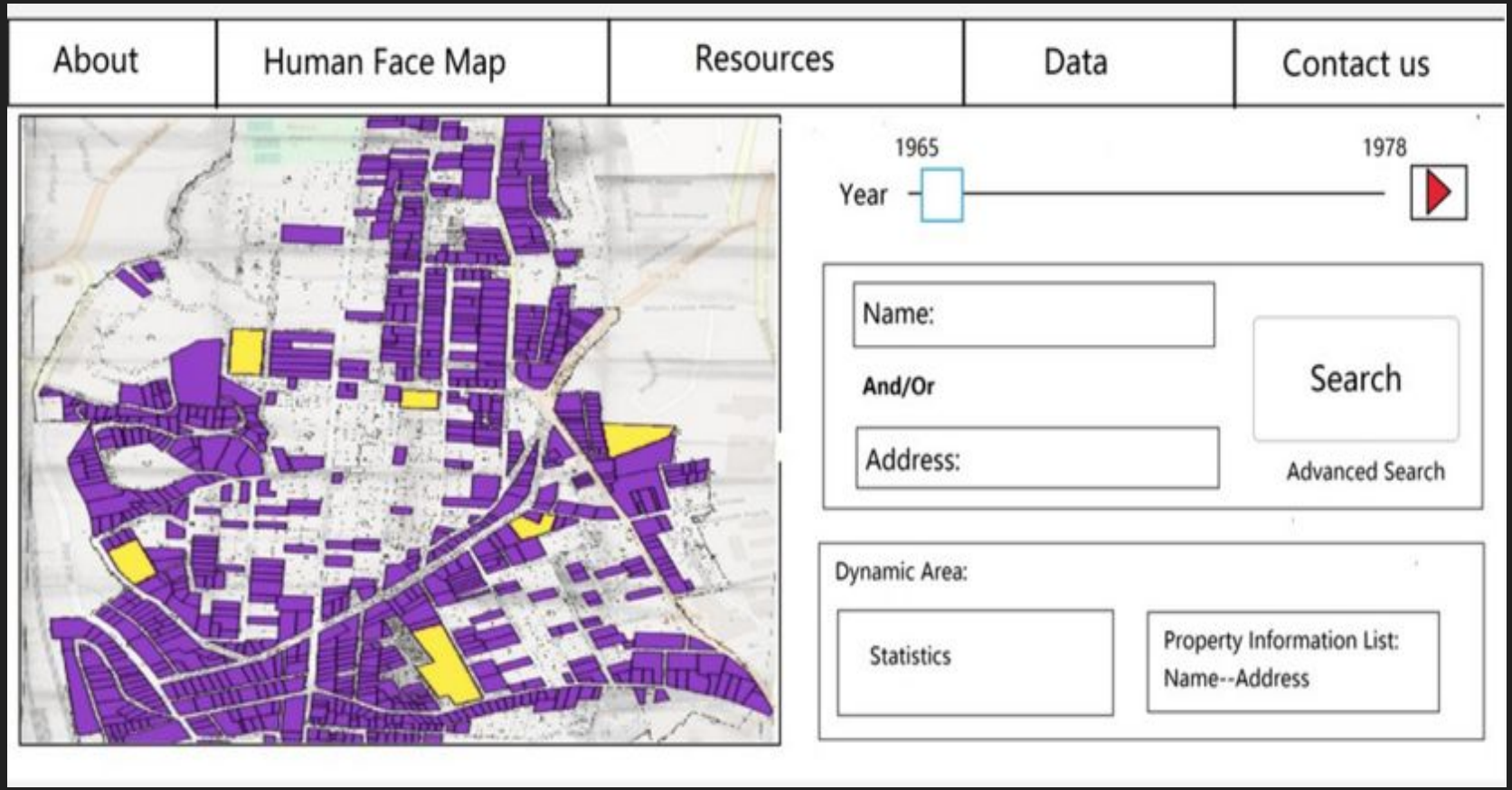
User Feedback

WWW Prototyping



Interface/DB Design (Wired-frame)

Human Face of Big Data

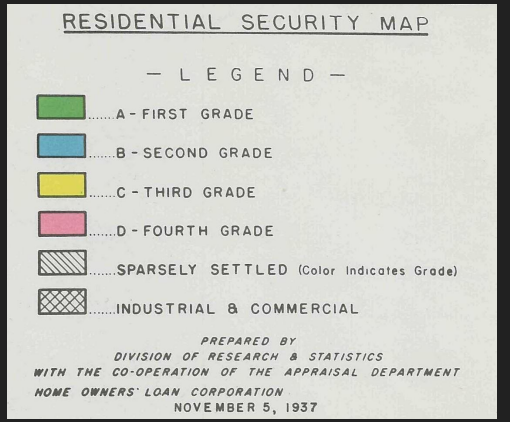


2. Mapping Inequality
[Racial Zoning]

UMD Student Team:
Mary Kendig
Myeong Lee
Sydney Vaile
Maddie Allen
Martin Almirón
Jhon De La Cruz
Shaina Destine
Erin Durham
Darlene Reyes
Benjamin Sagay
Richard Bool

Historical Context

- Home Owners Loan Corporation 1930's - 1940's
  - Rated neighborhoods by racial makeup
  - Areas without loans fell apart
- 1950's Urban Renewal targeted areas for clearance
- Result: Mass displacement
- RG195: Federal Home Loan Bank Board, HOLC,1933 - 1951
- Contains Maps, Neighborhood Surveys, Loan Information



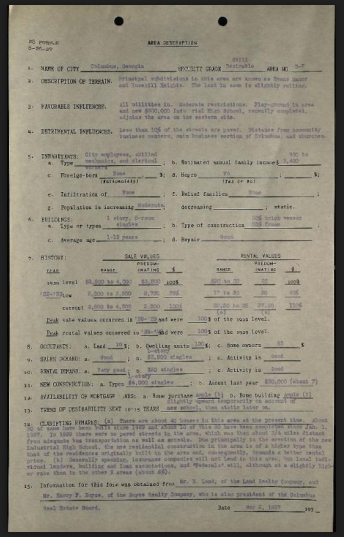
Mapping Inequality

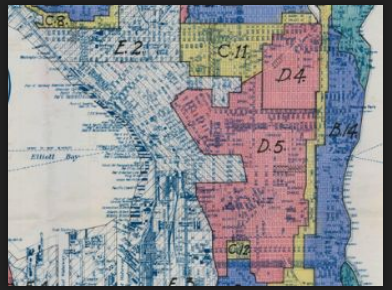
Documents

- Each survey corresponded to city map
  - Green: White / Wealthy = Best
  - Blue: White / Working = Still Desirable
  - Yellow: Foreign / Increase in PoC = Declining
  - Red: Black and Hispanic = Hazardous

Collection Statistics

- 150 Boxes
- Over 6,000 surveys alone
- 250 cities





AREA DESCRIPTION

Security Map of LOS ANGELES COUNTY

1. POPULATION: *a. Increasing* Slowly *Decreasing* *Static*

b. Class and Occupation Jewish professional & business men, Mexican laborers, WPA workers, etc. Income $700 to $2000 and up

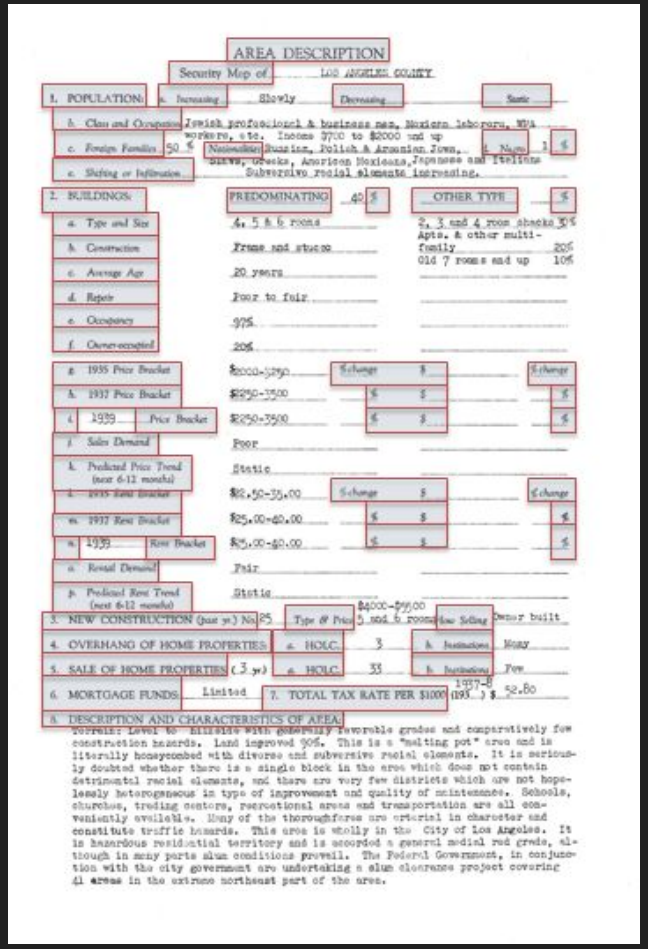
c. Foreign Families 50 % *Nationalities* Russian, Polish & Armenian Jews, Slavs, Greeks, American Mexicans, Japanese and Italians *d. Negro* 1 %

e. Shifting or Infiltration Subversive racial elements increasing.

8. DESCRIPTION AND CHARACTERISTICS OF AREA:

Terrain: Level to hillside with generally favorable grades and comparatively few construction hazards. Land improved 90%. This is a "melting pot" area and is literally honeycombed with diverse and subversive racial elements. It is seriously doubted whether there is a single block in the area which does not contain detrimental racial elements, and there are very few districts which are not hopelessly heterogeneous in type of improvement and quality of maintenance. Schools, churches, trading centers, recreational areas and transportation are all conveniently available. Many of the thoroughfares are arterial in character and constitute traffic hazards. This area is wholly in the City of Los Angeles. It is hazardous residential territory and is accorded a general medial red grade, although in many parts slum conditions prevail. The Federal Government, in conjunction with the city government are undertaking a slum clearance project covering 41 areas in the extreme northeast part of the area.

9. LOCATION Boyle Heights SECURITY GRADE 4th AREA NO. D-53 DATE 4-19-39



AREA DESCRIPTION

Security Map of LOS ANGELES COUNTY

1. POPULATION: a. Increasing — Slowly — Decreasing — Static

b. Class and Occupation: Jewish professional & business men, Mexican laborers, WPA workers, etc. Income $700 to $2000 and up

c. Foreign Families: 50 % Nationalities: Russian, Polish & Armenian Jews, Slavs, Greeks, American Mexicans, Japanese and Italians — d. Negro: 1 %

e. Shifting or Infiltration: Subversive racial elements increasing.

2. BUILDINGS:

	PREDOMINATING	%	OTHER TYPE	%
a. Type and Size	4, 5 & 6 rooms		2, 3 and 4 room shacks	30%
b. Construction	Frame and stucco		Apts. & other multi-family	20%
c. Average Age	20 years		Old 7 rooms and up	10%
d. Repair	Poor to fair			
e. Occupancy	90%			
f. Owner-occupied	20%			
g. 1935 Price Bracket	$2000-3250	% change $		% change
h. 1937 Price Bracket	$2250-3500	% $		%
i. 1939 Price Bracket	$2250-3500	% $		%
j. Sales Demand	Poor			
k. Predicted Price Trend (next 6-12 months)	Static			
l. 1935 Rent Bracket	$18.50-35.00	% change $		% change
m. 1937 Rent Bracket	$25.00-40.00	% $		%
n. 1939 Rent Bracket	$25.00-40.00	% $		%
o. Rental Demand	Fair			
p. Predicted Rent Trend (next 6-12 months)	Static			

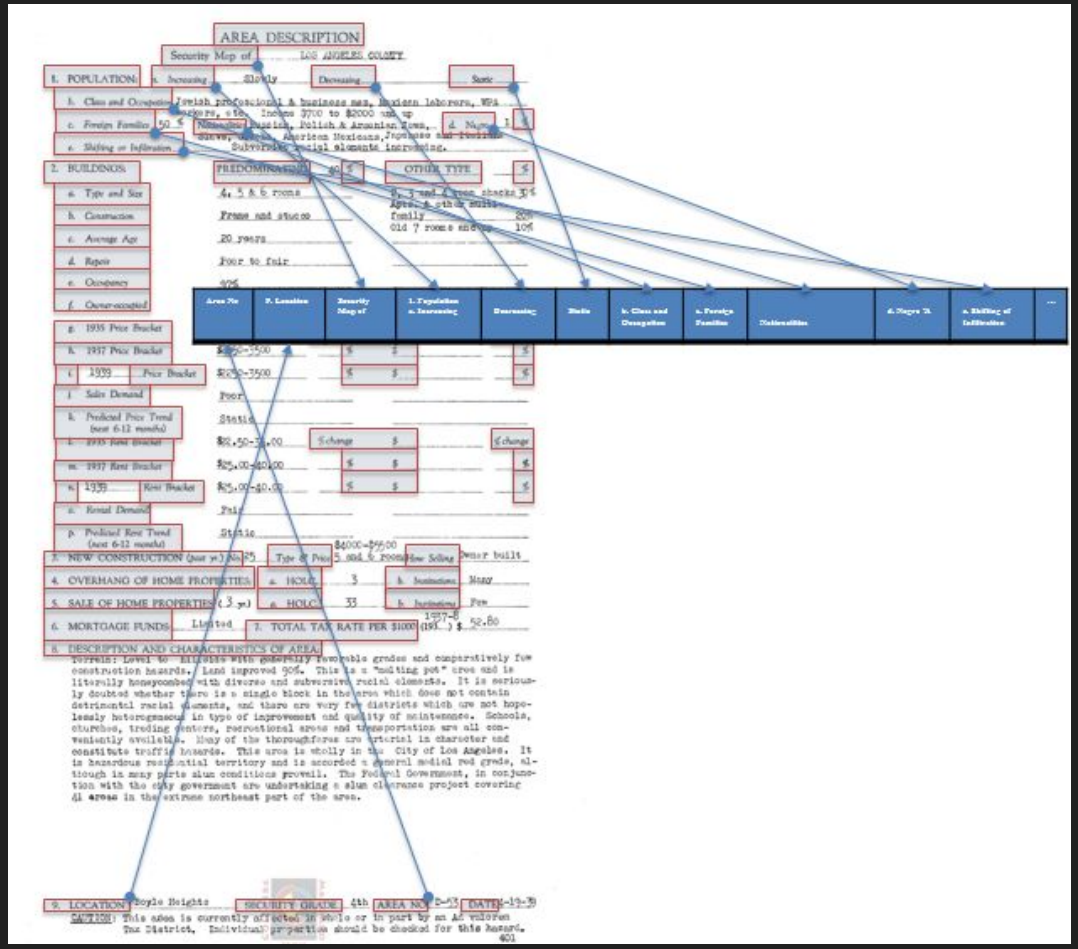
3. NEW CONSTRUCTION (last yr.) No. 85 — Type & Price: $4000-$5500, 5 and 6 rooms — How Selling: Owner built

4. OVERHANG OF HOME PROPERTIES: a. HOLC: 3 — b. Institutions: Many

5. SALE OF HOME PROPERTIES (3 yr.): a. HOLC: 33 — b. Institutions: Few

6. MORTGAGE FUNDS: Limited — 7. TOTAL TAX RATE PER $1000 (1937-8 (193) $ 52.80

8. DESCRIPTION AND CHARACTERISTICS OF AREA:
Terrain: Level to Hilside with generally favorable grades and comparatively few construction hazards. Land improved 90%. This is a "melting pot" area and is literally honeycombed with diverse and subversive racial elements. It is seriously doubted whether there is a single block in the area which does not contain detrimental racial elements, and there are very few districts which are not hopelessly heterogeneous in type of improvement and quality of maintenance. Schools, churches, trading centers, recreational areas and transportation are all conveniently available. Many of the thoroughfares are arterial in character and constitute traffic hazards. This area is wholly in the City of Los Angeles. It is hazardous residential territory and is accorded a general medial red grade, although in many parts slum conditions prevail. The Federal Government, in conjunction with the city government are undertaking a slum clearance project covering 41 areas in the extreme northeast part of the area.



AREA DESCRIPTION

Security Map of LOS ANGELES COUNTY

1. POPULATION: a. Increasing — Slowly — Decreasing — Static

b. Class and Occupation: Jewish professional & business men, Mexican laborers, WPA workers, etc. Income $700 to $2000 and up

c. Foreign Families: 50 % Nationalities: Russian, Polish & Armenian Jews, Slavs, Greeks, American Mexicans, Japanese and Italians — d. Negro: 1 %

e. Shifting or Infiltration: Subversive racial elements increasing.

2. BUILDINGS:

	PREDOMINATING	%	OTHER TYPE	%
a. Type and Size	4, 5 & 6 rooms		2, 3 and 4 room shacks	30%
b. Construction	Frame and stucco		Apts. & other multi-family	20%
c. Average Age	20 years		Old 7 rooms and up	10%
d. Repair	Poor to fair			
e. Occupancy	90%			
f. Owner-occupied	20%			
g. 1935 Price Bracket	$2000-3250	% change $		% change
h. 1937 Price Bracket	$2250-3500	% $		%
i. 1939 Price Bracket	$2250-3500	% $		%
j. Sales Demand	Poor			
k. Predicted Price Trend (next 6-12 months)	Static			
l. 1935 Rent Bracket	$18.50-35.00	% change $		% change
m. 1937 Rent Bracket	$25.00-40.00	% $		%
n. 1939 Rent Bracket	$25.00-40.00	% $		%
o. Rental Demand	Fair			
p. Predicted Rent Trend (next 6-12 months)	Static			

3. NEW CONSTRUCTION (last yr.) No. 85 — Type & Price: $4000-$5500, 5 and 6 rooms — How Selling: Owner built

4. OVERHANG OF HOME PROPERTIES: a. HOLC: 3 — b. Institutions: Many

5. SALE OF HOME PROPERTIES (3 yr.): a. HOLC: 33 — b. Institutions: Few

6. MORTGAGE FUNDS: Limited — 7. TOTAL TAX RATE PER $1000 (1937-8 (193) $ 52.80

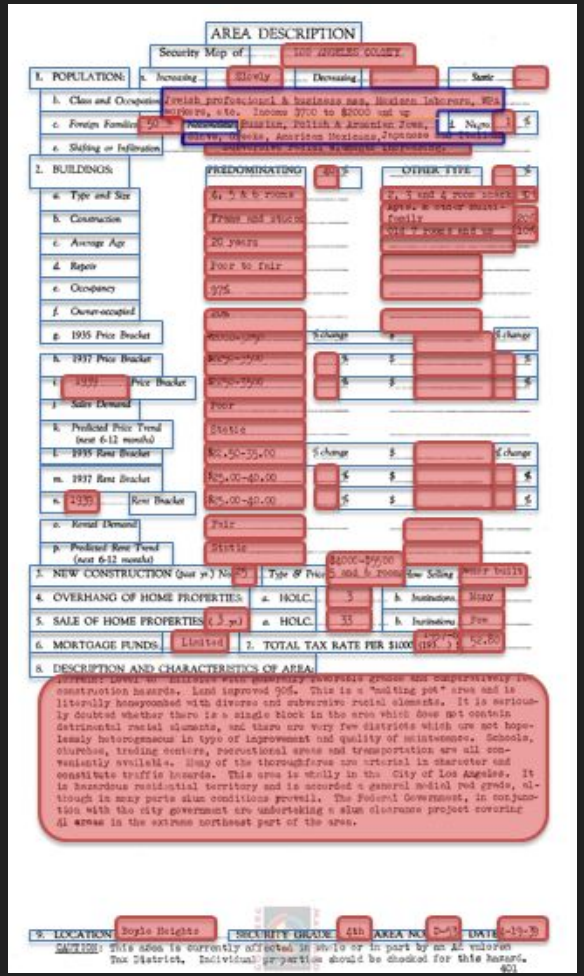
8. DESCRIPTION AND CHARACTERISTICS OF AREA:
Terrain: Level to Hilside with generally favorable grades and comparatively few construction hazards. Land improved 90%. This is a "melting pot" area and is literally honeycombed with diverse and subversive racial elements. It is seriously doubted whether there is a single block in the area which does not contain detrimental racial elements, and there are very few districts which are not hopelessly heterogeneous in type of improvement and quality of maintenance. Schools, churches, trading centers, recreational areas and transportation are all conveniently available. Many of the thoroughfares are arterial in character and constitute traffic hazards. This area is wholly in the City of Los Angeles. It is hazardous residential territory and is accorded a general medial red grade, although in many parts slum conditions prevail. The Federal Government, in conjunction with the city government are undertaking a slum clearance project covering 41 areas in the extreme northeast part of the area.

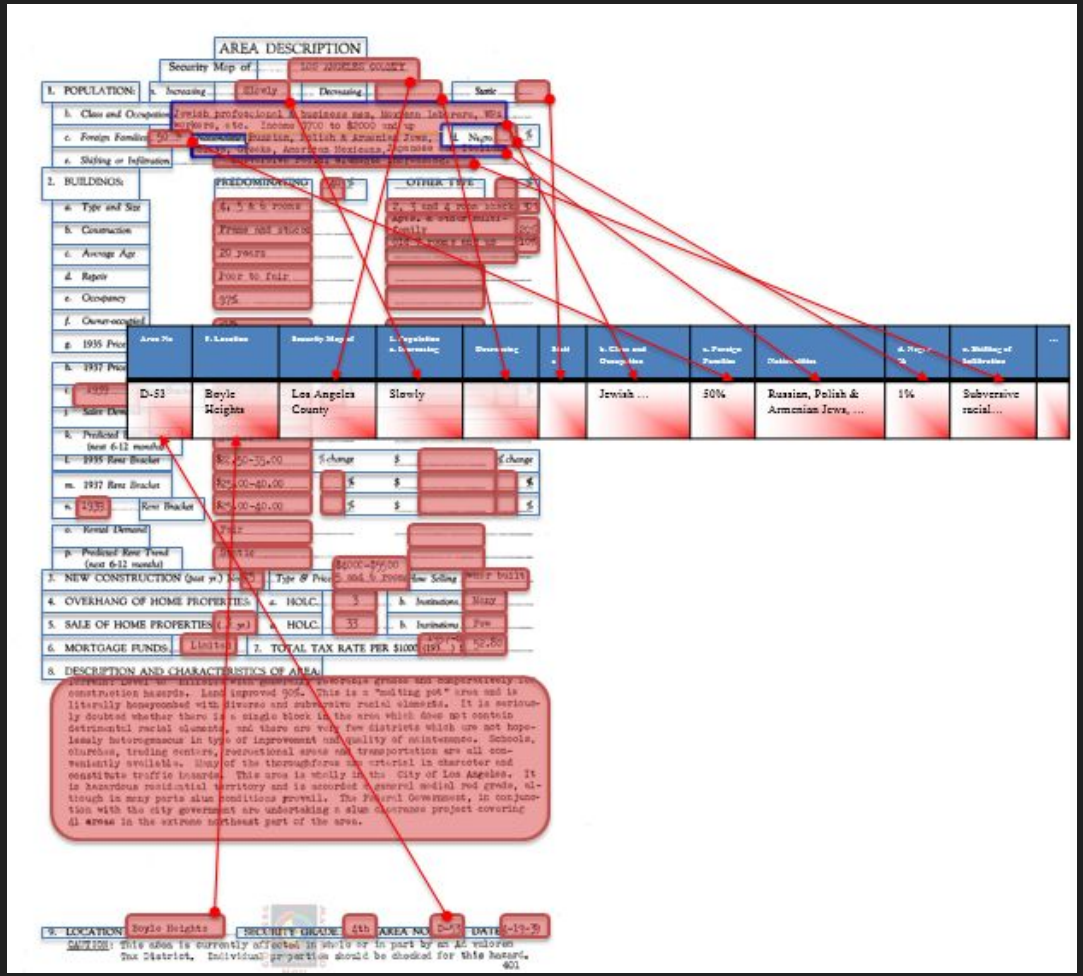
9. LOCATION: Temple Heights — SECURITY GRADE: 4th — AREA NO. D-55 — DATE: 1-19-39

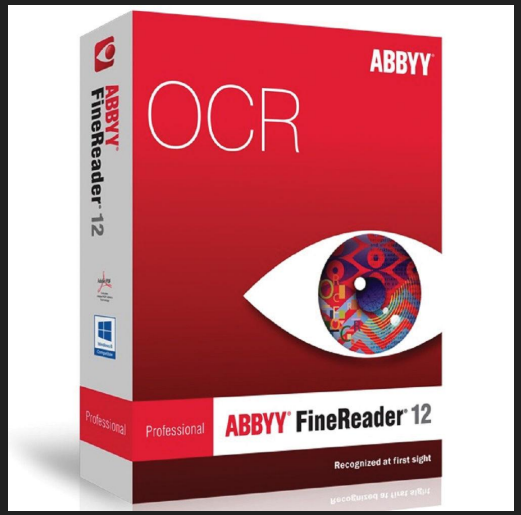
SECTION: This area is currently affected in whole or in part by an Ad valorem Tax District. Individual properties should be checked for this hazard.

401

Area No. | F. Location | Security Map of | 1. Population a. Increasing | Decreasing | Static | b. Class and Occupation | c. Foreign Families | Nationalities | d. Negro % | e. Shifting or Infiltration | ...


AREA DESCRIPTION
Security Map of LOS ANGELES COUNTY
1. POPULATION: a. Increasing Slowly Decreasing Static
b. Class and Occupation Jewish professional & business men, Mexican laborers, WPA workers, etc. Income $700 to $2000 and up
c. Foreign Families 50% Predominating Russian, Polish & Armenian Jews, Chinese, Greeks, American Mexicans, Japanese
d. Negro 1 %
e. Shifting or Infiltration Subversive racial elements...
2. BUILDINGS: PREDOMINATING 80 % OTHER TYPE 20 %
a. Type and Size 4, 5 & 6 rooms 2, 3 and 4 room stock 30%
b. Construction Frame and stucco
c. Average Age 20 years
d. Repair Poor to fair
e. Occupancy 97%
f. Owner-occupied 20%
g. 1935 Price Bracket $2,750-3,500 % change $ $2,250-2,750 % change
h. 1937 Price Bracket $2,750-3,500 % change $ $2,250-2,750 %
i. 1939 Price Bracket $2,750-3,500 % change $ $2,250-2,750 %
j. Sales Demand Poor
k. Predicted Price Trend (next 6-12 months) Static
l. 1935 Rent Bracket $35.00-40.00 % change $ % change
m. 1937 Rent Bracket $35.00-40.00 % $ $ %
n. 1939 Rent Bracket $35.00-40.00 % $ $ %
o. Rental Demand Fair
p. Predicted Rent Trend (next 6-12 months) Static
3. NEW CONSTRUCTION (last yr.) No. 25 Type & Price $4000-$5000 5 and 6 room Now Selling Mostly built
4. OVERHANG OF HOME PROPERTIES a. HOLC 3 b. Institutions None
5. SALE OF HOME PROPERTIES (last 3 yrs.) a. HOLC 33 b. Institutions Few
6. MORTGAGE FUNDS Limited 7. TOTAL TAX RATE PER $100 (1937) $2.80
8. DESCRIPTION AND CHARACTERISTICS OF AREA:
Different level in Hillside with generally favorable grades and comparatively few construction hazards. Land improved 90%. This is a "melting pot" area and is literally honeycombed with diverse and subversive racial elements. It is seriously doubted whether there is a single block in the area which does not contain detrimental racial elements, and there are very few districts which are not hopelessly heterogeneous in type of improvement and quality of maintenance. Schools, churches, trading centers, recreational areas and transportation are all conveniently available. Many of the thoroughfares are arterial in character and constitute traffic hazards. This area is wholly in the City of Los Angeles. It is hazardous residential territory and is accorded a general medial red grade, although in many parts slum conditions prevail. The Federal Government, in conjunction with the city government are undertaking a slum clearance project covering all areas in the extreme northeast part of the area.
9. LOCATION Boyle Heights SECURITY GRADE 4th AREA NO. D-53 DATE 6-19-39
CAUTION: This area is currently affected in whole or in part by an AZ valorem Tax District. Individual properties should be checked for this hazard.
401

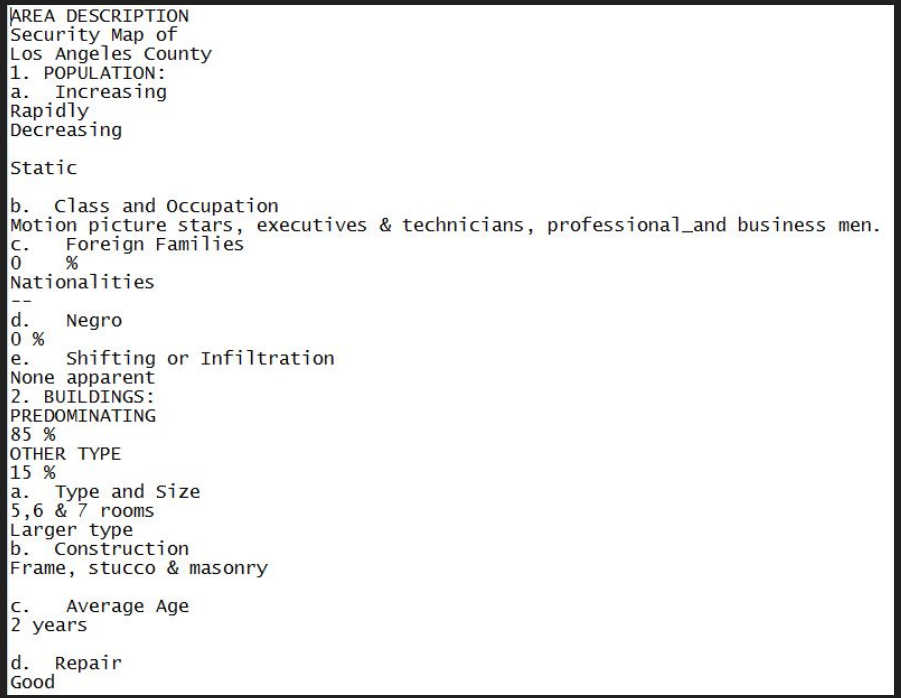

AREA DESCRIPTION
Security Map of LOS ANGELES COUNTY
1. POPULATION: a. Increasing Slowly Decreasing Static
b. Class and Occupation Jewish professional & business men, Mexican laborers, WPA workers, etc. Income $700 to $2000 and up
c. Foreign Families 50% Predominating Russian, Polish & Armenian Jews, Chinese, Greeks, American Mexicans, Japanese
d. Negro 1 %
e. Shifting or Infiltration Subversive racial elements...
2. BUILDINGS: PREDOMINATING 80 % OTHER TYPE 20 %
a. Type and Size 4, 5 & 6 rooms 2, 3 and 4 room stock 30%
b. Construction Frame and stucco
c. Average Age 20 years
d. Repair Poor to fair
e. Occupancy 97%
f. Owner-occupied 20%
g. 1935 Price Bracket $2,750-3,500 % change $ $2,250-2,750 % change
h. 1937 Price Bracket $2,750-3,500 % change $ $2,250-2,750 %
i. 1939 Price Bracket $2,750-3,500 % change $ $2,250-2,750 %
j. Sales Demand Poor
k. Predicted Price Trend (next 6-12 months) Static
l. 1935 Rent Bracket $35.00-40.00 % change $ % change
m. 1937 Rent Bracket $35.00-40.00 % $ $ %
n. 1939 Rent Bracket $35.00-40.00 % $ $ %
o. Rental Demand Fair
p. Predicted Rent Trend (next 6-12 months) Static
3. NEW CONSTRUCTION (last yr.) No. 25 Type & Price $4000-$5000 5 and 6 room Now Selling Mostly built
4. OVERHANG OF HOME PROPERTIES a. HOLC 3 b. Institutions None
5. SALE OF HOME PROPERTIES (last 3 yrs.) a. HOLC 33 b. Institutions Few
6. MORTGAGE FUNDS Limited 7. TOTAL TAX RATE PER $100 (1937) $2.80
8. DESCRIPTION AND CHARACTERISTICS OF AREA:
Different level in Hillside with generally favorable grades and comparatively few construction hazards. Land improved 90%. This is a "melting pot" area and is literally honeycombed with diverse and subversive racial elements. It is seriously doubted whether there is a single block in the area which does not contain detrimental racial elements, and there are very few districts which are not hopelessly heterogeneous in type of improvement and quality of maintenance. Schools, churches, trading centers, recreational areas and transportation are all conveniently available. Many of the thoroughfares are arterial in character and constitute traffic hazards. This area is wholly in the City of Los Angeles. It is hazardous residential territory and is accorded a general medial red grade, although in many parts slum conditions prevail. The Federal Government, in conjunction with the city government are undertaking a slum clearance project covering all areas in the extreme northeast part of the area.
9. LOCATION Boyle Heights SECURITY GRADE 4th AREA NO. D-53 DATE 6-19-39
CAUTION: This area is currently affected in whole or in part by an AZ valorem Tax District. Individual properties should be checked for this hazard.
401
Area No. | 1. Location | Security Map of | 2. a. Population b. Increasing | Decreasing | Static | b. Class and Occupation | c. Foreign Families | Predominating | d. Negro % | e. Shifting or Infiltration | ...
D-53 | Boyle Heights | Los Angeles County | Slowly | | | Jewish ... | 50% | Russian, Polish & Armenian Jews, ... | 1% | Subversive racial... |



ABBYY FineReader OCR Software

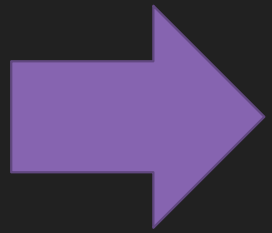
```
AREA DESCRIPTION
Security Map of
Los Angeles County
1. POPULATION:
a.  Increasing
Rapidly
Decreasing

Static

b.  Class and Occupation
Motion picture stars, executives & technicians, professional_and business men.        Income $3600-$6000 & up.
c.   Foreign Families
0    %
Nationalities
--
d.   Negro
0 %
e.   Shifting or Infiltration
None apparent
2. BUILDINGS:
PREDOMINATING
85 %
OTHER TYPE
15 %
a.  Type and Size
5,6 & 7 rooms
Larger type
b.  Construction
Frame, stucco & masonry

c.   Average Age
2 years

d.  Repair
Good
```

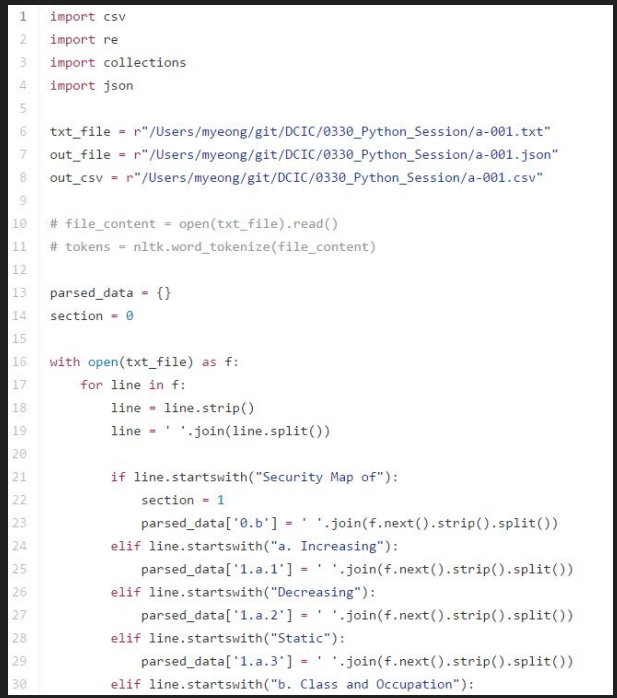
Plain .TXT File



```
AREA DESCRIPTION
Security Map of
Los Angeles County
1. POPULATION:
a.  Increasing
Rapidly
Decreasing

Static

b.  Class and Occupation
Motion picture stars, executives & technicians, professional and business men.
c.   Foreign Families
0    %
Nationalities
--
d.   Negro
0 %
e.   Shifting or Infiltration
None apparent
2. BUILDINGS:
PREDOMINATING
85 %
OTHER TYPE
15 %
a.  Type and Size
5,6 & 7 rooms
Larger type
b.  Construction
Frame, stucco & masonry

c.   Average Age
2 years

d.  Repair
Good
```

Plain .TXT File



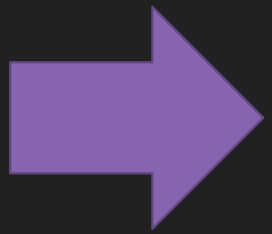


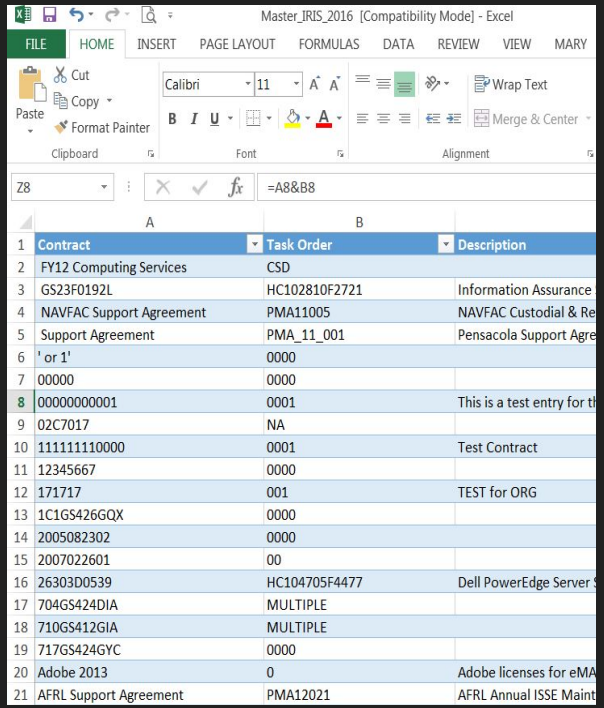
```python
import csv
import re
import collections
import json

txt_file = r"/Users/myeong/git/DCIC/0330_Python_Session/a-001.txt"
out_file = r"/Users/myeong/git/DCIC/0330_Python_Session/a-001.json"
out_csv = r"/Users/myeong/git/DCIC/0330_Python_Session/a-001.csv"

# file_content = open(txt_file).read()
# tokens = nltk.word_tokenize(file_content)

parsed_data = {}
section = 0

with open(txt_file) as f:
    for line in f:
        line = line.strip()
        line = ' '.join(line.split())

        if line.startswith("Security Map of"):
            section = 1
            parsed_data['0.b'] = ' '.join(f.next().strip().split())
        elif line.startswith("a. Increasing"):
            parsed_data['1.a.1'] = ' '.join(f.next().strip().split())
        elif line.startswith("Decreasing"):
            parsed_data['1.a.2'] = ' '.join(f.next().strip().split())
        elif line.startswith("Static"):
            parsed_data['1.a.3'] = ' '.join(f.next().strip().split())
        elif line.startswith("b. Class and Occupation"):
```

Python Programming Script

```python
import csv
import re
import collections
import json

txt_file = r"/Users/myeong/git/DCIC/0330_Python_Session/a-001.txt"
out_file = r"/Users/myeong/git/DCIC/0330_Python_Session/a-001.json"
out_csv = r"/Users/myeong/git/DCIC/0330_Python_Session/a-001.csv"

# file_content = open(txt_file).read()
# tokens = nltk.word_tokenize(file_content)

parsed_data = {}
section = 0

with open(txt_file) as f:
    for line in f:
        line = line.strip()
        line = ' '.join(line.split())

        if line.startswith("Security Map of"):
            section = 1
            parsed_data['0.b'] = ' '.join(f.next().strip().split())
        elif line.startswith("a. Increasing"):
            parsed_data['1.a.1'] = ' '.join(f.next().strip().split())
        elif line.startswith("Decreasing"):
            parsed_data['1.a.2'] = ' '.join(f.next().strip().split())
        elif line.startswith("Static"):
            parsed_data['1.a.3'] = ' '.join(f.next().strip().split())
        elif line.startswith("b. Class and Occupation"):
```

Python Programming Script

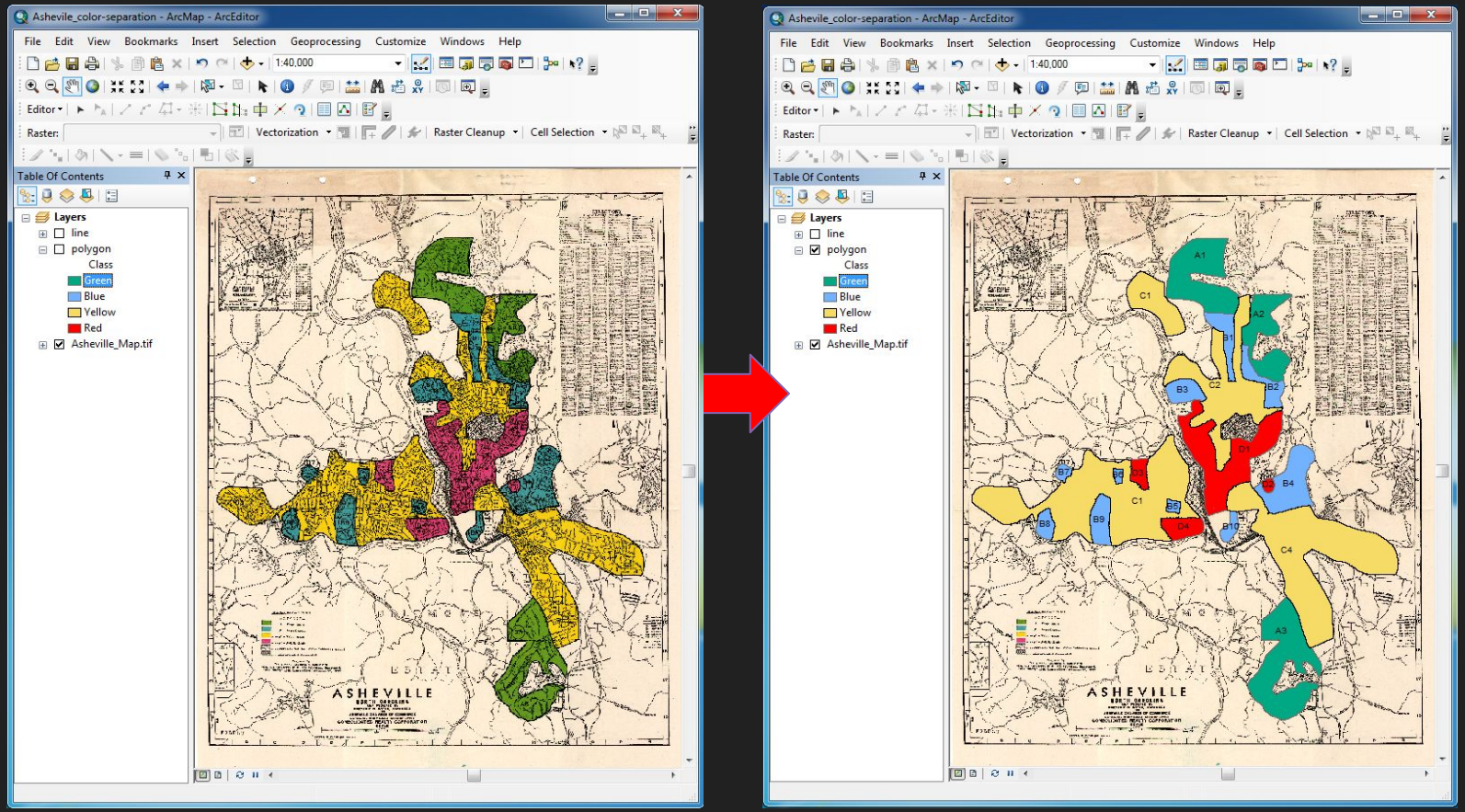




	A	B	
1	Contract	Task Order	Description
2	FY12 Computing Services	CSD	
3	GS23F0192L	HC102810F2721	Information Assurance
4	NAVFAC Support Agreement	PMA11005	NAVFAC Custodial & Re
5	Support Agreement	PMA_11_001	Pensacola Support Agre
6	' or 1'	0000	
7	00000	0000	
8	00000000001	0001	This is a test entry for t
9	02C7017	NA	
10	111111110000	0001	Test Contract
11	12345667	0000	
12	171717	001	TEST for ORG
13	1C1GS426GQX	0000	
14	2005082302	0000	
15	2007022601	00	
16	26303D0539	HC104705F4477	Dell PowerEdge Server
17	704GS424DIA	MULTIPLE	
18	710GS412GIA	MULTIPLE	
19	717GS424GYC	0000	
20	Adobe 2013	0	Adobe licenses for eMA
21	AFRL Support Agreement	PMA12021	AFRL Annual ISSE Maint

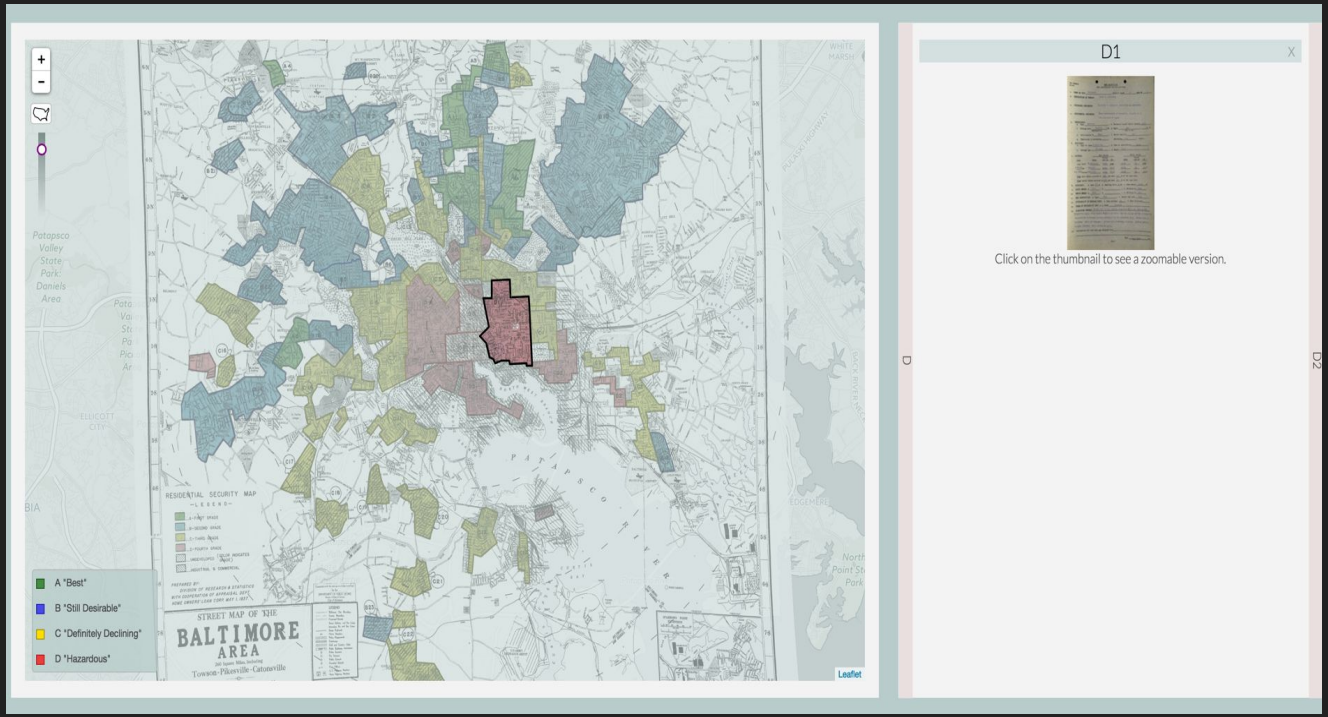
Excel Spreadsheet

Geo - Tracing


Ashevile_color-separation - ArcMap - ArcEditor
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
1:40,000
Editor
Raster:
Vectorization
Raster Cleanup
Cell Selection
Table Of Contents
Layers
line
polygon
Class
Green
Blue
Yellow
Red
Asheville_Map.tif
ASHEVILLE
NORTH CAROLINA

Mapping Inequality US

Link: https://dsl.richmond.edu/panorama/redlining/



Partner Universities:

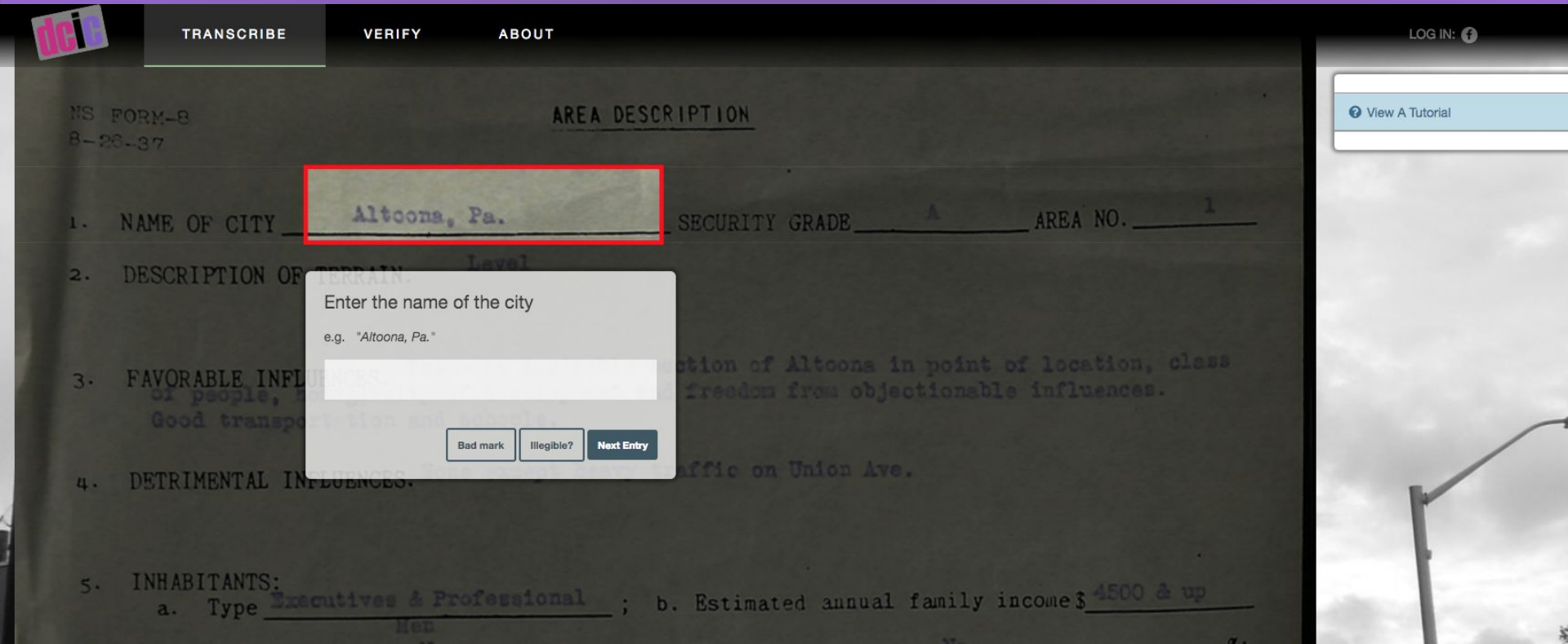
U. Maryland's iSchool's DCIC Center
Dr. Richard Marciano

U. of Richmond's Digital Scholarship Lab
Dr. Robert Nelson

Virginia Tech
Dr. LaDale Winling

Johns Hopkins University
Dr. N.D.B. Connolly

SCRIBE: Crowdsourcing Platform



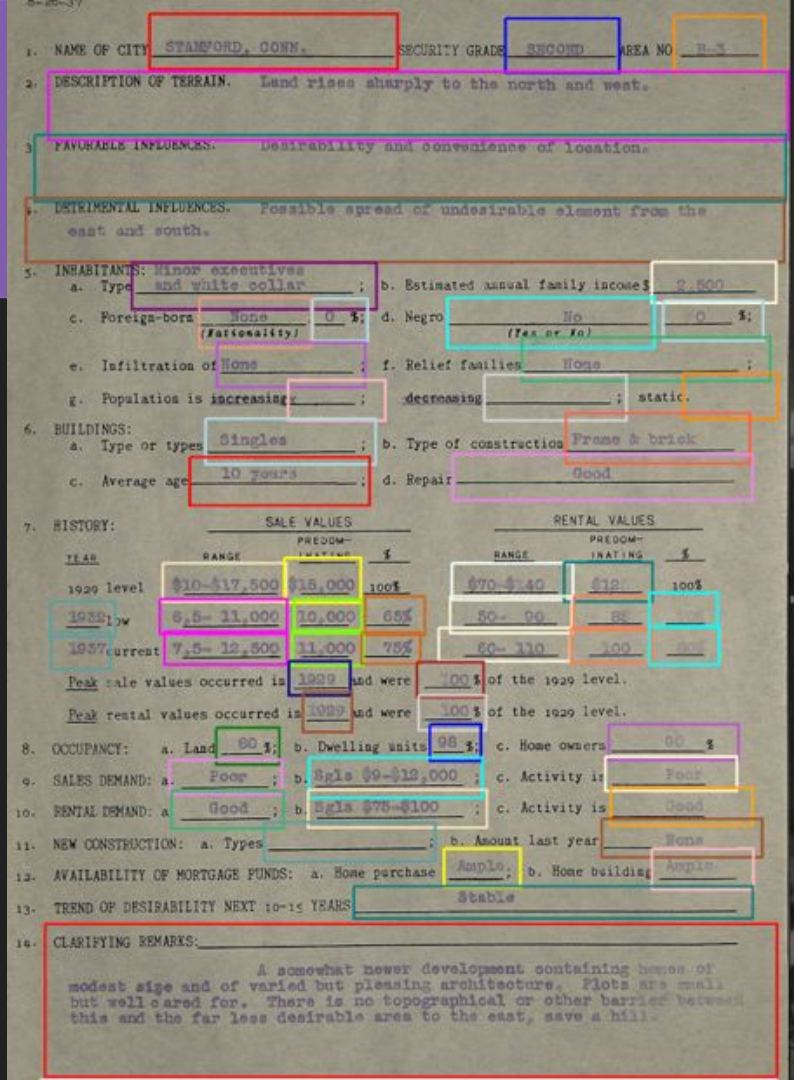
SCRIBE: Crowdsourcing Platform

Originally Developed by NYPL (https://github.com/zooniverse/scribeAPI)

Open Source Software written in Ruby on Rails

Collaborative Transcription and Verification of Texts that cannot be read automatically: CrowdSourcing

Automatic Generation of Marking Boxes



3. St. Louis Voyage

[Refugee Narratives]

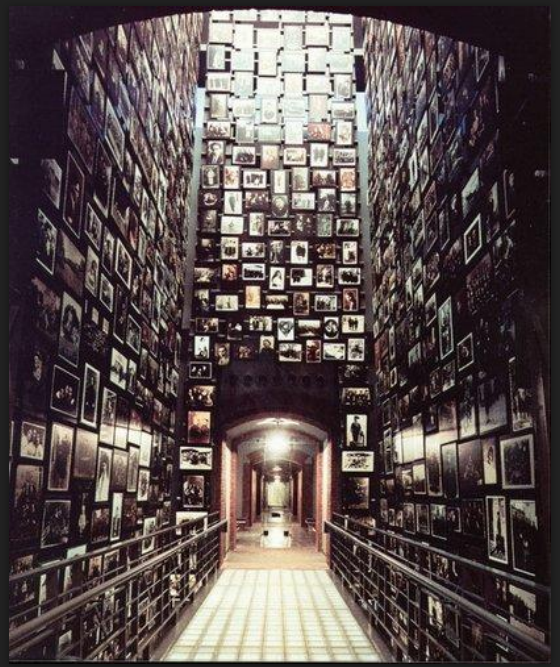
UMD Student Team:
Mary Kendig
Myuresh Amdekar
Brian Refford
Sohan Shah
Pal Doshi
Yuting Liao
Ruchira Kapoor
Sohan Shah

3. St. Louis Voyage
[Refugee Narratives]

- May 13, 1939 German ship SS St. Louis sailed 937 Jewish refugees from Third Reich to Havana, Cuba
- Despite arriving, SS St. Louis turned away
- Captain sailed to US
- State Department and White House denied entry

St. Louis Voyage

- United States Holocaust Memorial Museum determined fate of all 937 passengers
- Developed spreadsheets containing passenger data, events, and countries
- Continuously collecting photographs, stories, and documents



St. Louis Voyage: Building Archival Databases

- Students working directly with USHMM staff to conduct requirements gathering
- Using SQL programming and MySQL Workbench
- Building application for users to interact with data and collections

St. Louis Voyage: Data Analytics

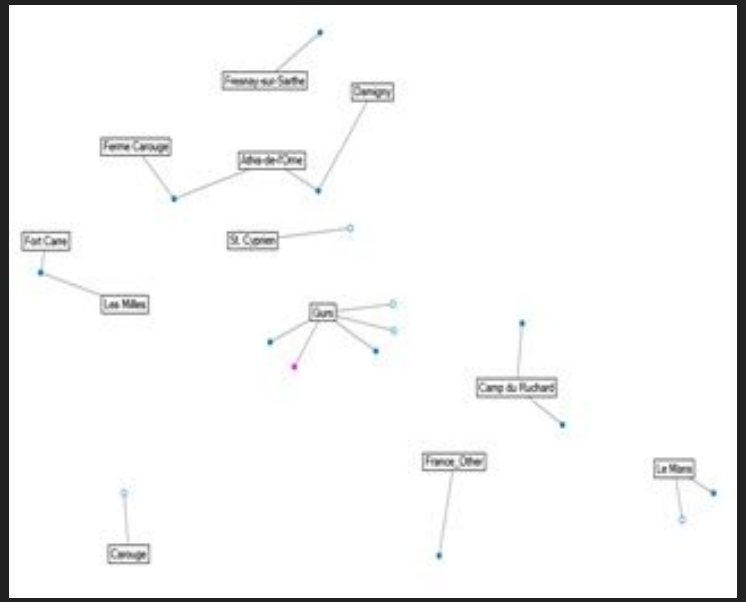
- Using R Programming to conduct data analytics
- Working with visualization software Tableau, NodeXL, Gephi
- Attributing coordinates to camps and cities to visualize movement of families

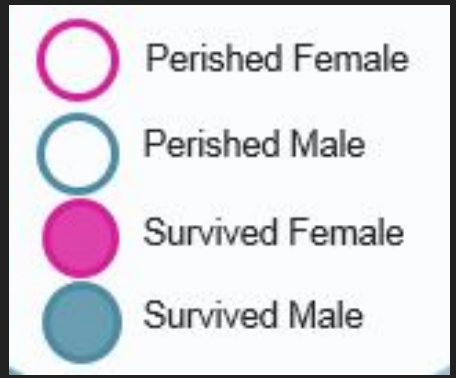
Network of People in Internment

Event Type: *Interment* ***Country****: France*

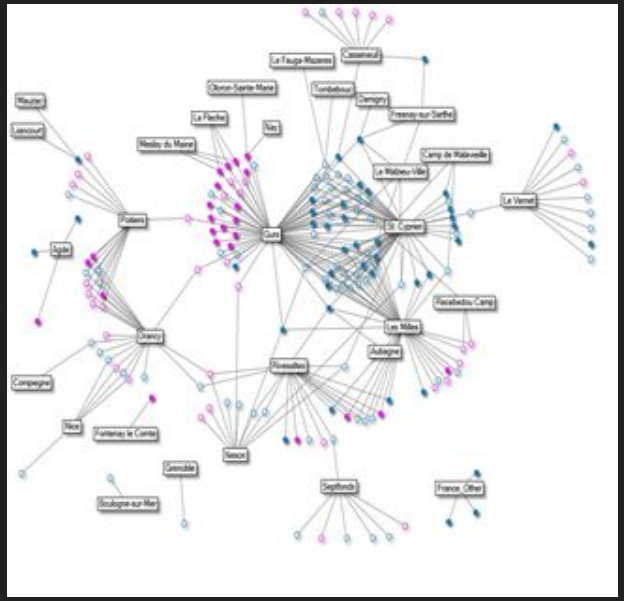
By Yuting Lao

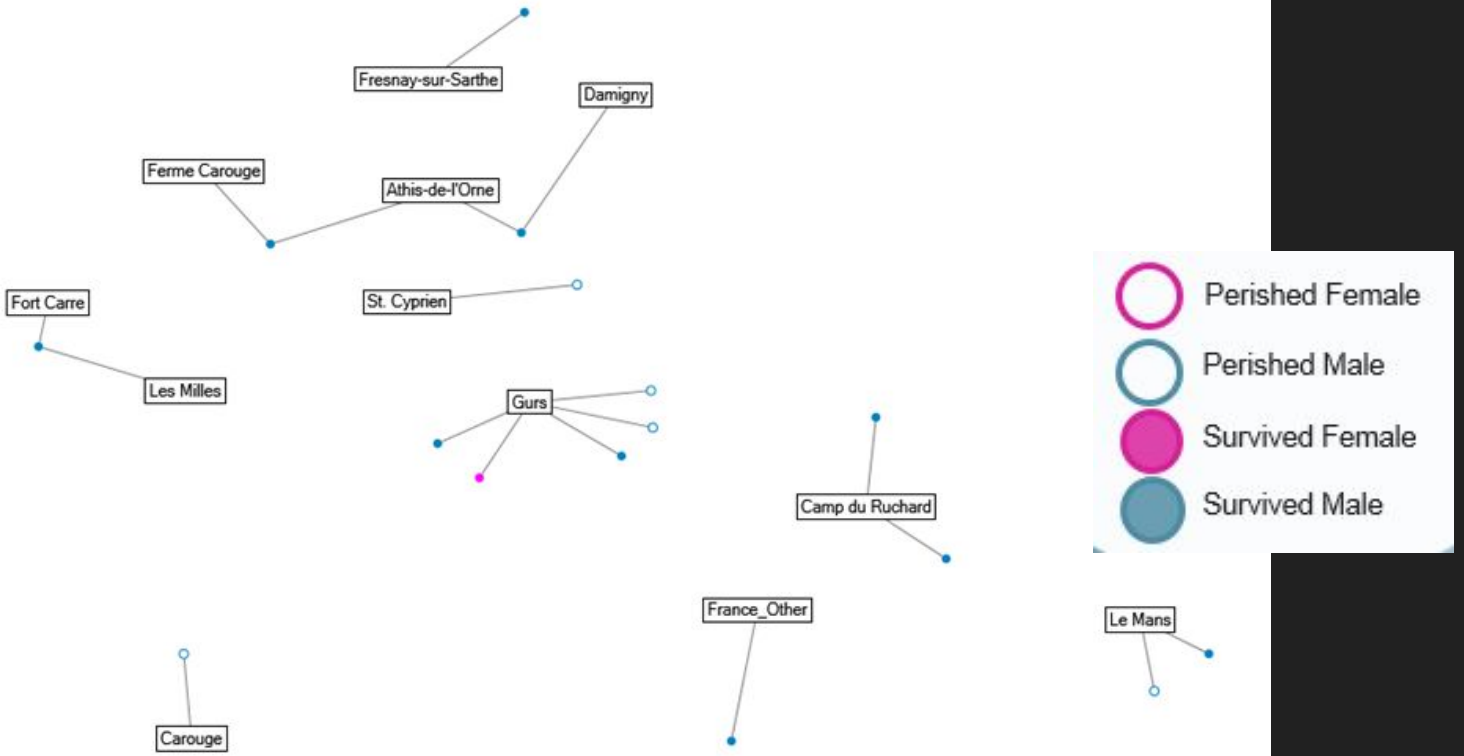
- *1939–40 (before Nazi invasion)*

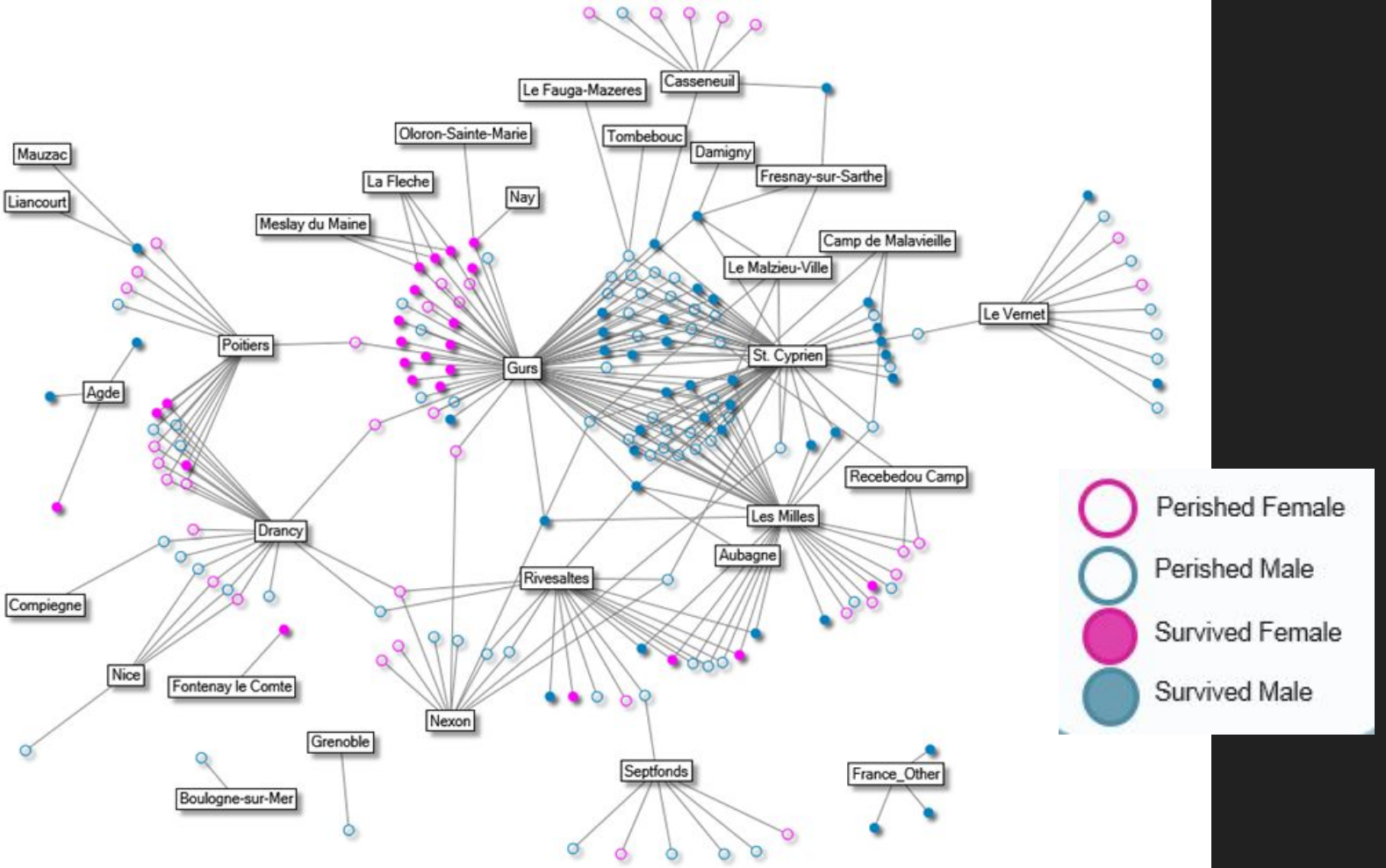




1940-46 (after Nazi invasion)




Fresnay-sur-Sarthe
Damigny
Ferme Carouge
Athis-de-l'Orne
Fort Carre
St. Cyprien
Les Milles
Gurs
Camp du Ruchard
France_Other
Le Mans
Carouge
Perished Female
Perished Male
Survived Female
Survived Male


Casseneuil
Le Fauga-Mazeres
Oloron-Sainte-Marie
Tombebouc
Damigny
Fresnay-sur-Sarthe
Mauzac
La Fleche
Nay
Liancourt
Meslay du Maine
Camp de Malavieille
Le Malzieu-Ville
Le Vernet
Poitiers
Gurs
St. Cyprien
Agde
Recebedou Camp
Les Milles
Drancy
Aubagne
Rivesaltes
Compiegne
Nice
Fontenay le Comte
Nexon
Grenoble
Septfonds
France_Other
Boulogne-sur-Mer
Perished Female
Perished Male
Survived Female
Survived Male

4. WWII Japanese-American Camps

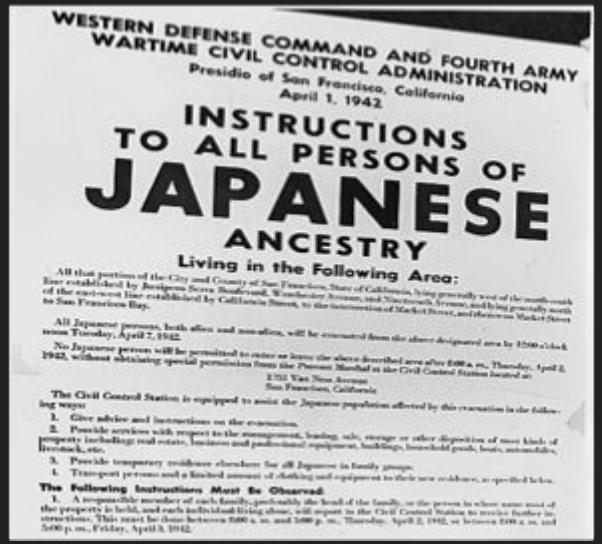
[Citizen Internment]

UMD Student Team:
Jiahui Wu
Kelsey Diemand
Meaghan Wilson
Vinita Atre
Diane Travis

UMD Faculty:
Richard Marciano
Bill Underwood

4. WWII Japanese-American Camps

[Citizen Internment]



WESTERN DEFENSE COMMAND AND FOURTH ARMY
WARTIME CIVIL CONTROL ADMINISTRATION
Presidio of San Francisco, California
April 1, 1942

INSTRUCTIONS TO ALL PERSONS OF JAPANESE ANCESTRY

Living in the Following Area:

All that portion of the City and County of San Francisco, State of California, lying generally west of the north-south line established by Junipero Serra Boulevard, Winchester Avenue, and Nineteenth Avenue, and lying generally north of the east-west line established by California Street, to the intersection of Market Street, and thence on Market Street to San Francisco Bay.

All Japanese persons, both alien and non-alien, will be evacuated from the above designated area by 12:00 o'clock noon Tuesday, April 7, 1942.

No Japanese person will be permitted to enter or leave the above-described area after 8:00 a. m., Thursday, April 2, 1942, without obtaining special permission from the Provost Marshal at the Civil Control Station located at:

1701 Van Ness Avenue
San Francisco, California

The Civil Control Station is equipped to assist the Japanese population affected by this evacuation in the following ways:

1. Give advice and instructions on the evacuation.
2. Provide services with respect to the management, leasing, sale, storage or other disposition of most kinds of property, including: real estate, business and professional equipment, buildings, household goods, boats, automobiles, livestock, etc.
3. Provide temporary residence elsewhere for all Japanese in family groups.
4. Transport persons and a limited amount of clothing and equipment to their new residence, as specified below.

The Following Instructions Must Be Observed:

1. A responsible member of each family, preferably the head of the family, or the person in whose name most of the property is held, and each individual living alone, will report to the Civil Control Station to receive further instructions. This must be done between 8:00 a. m. and 5:00 p. m., Thursday, April 2, 1942, or between 8:00 a. m. and 5:00 p. m., Friday, April 3, 1942.

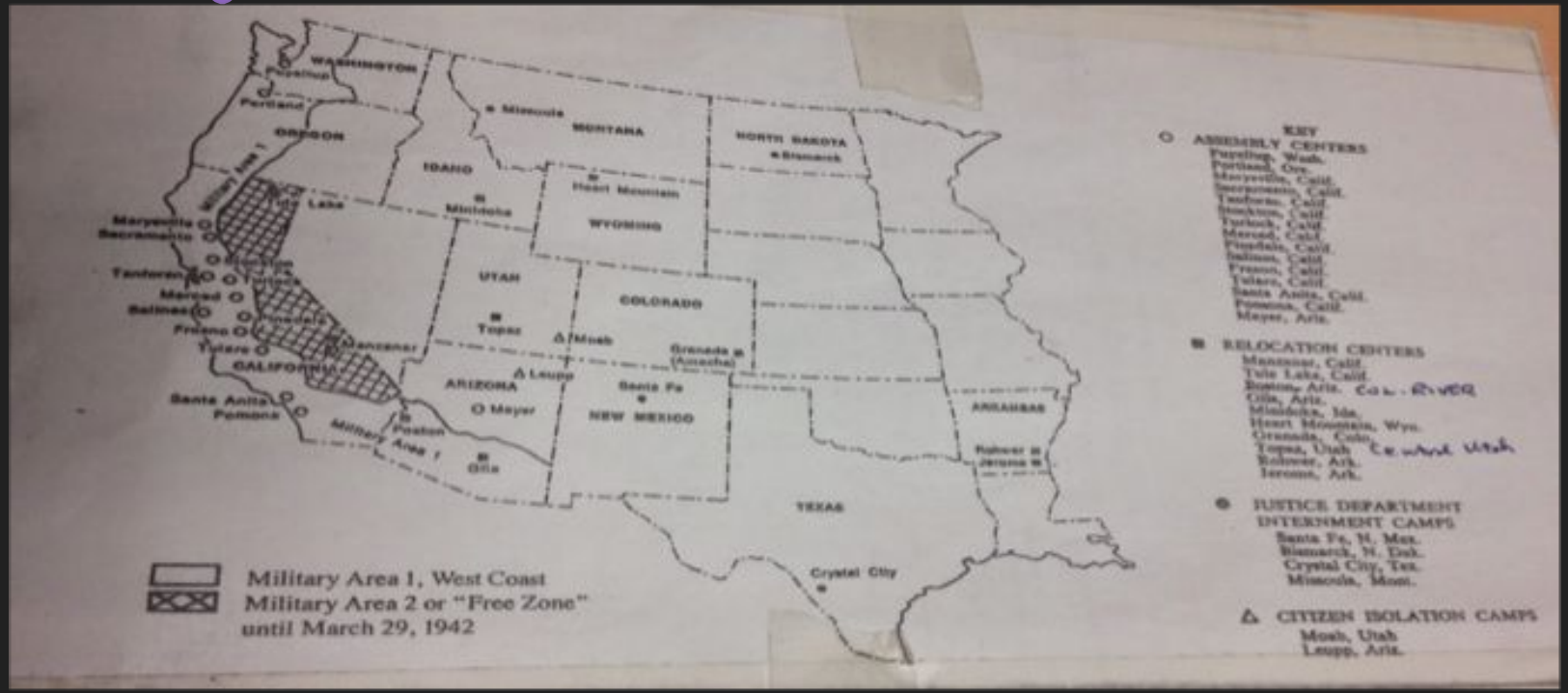
AETN's "In Their Words" project re: the Rohwer Internment Camp:

- Four-year old George Takei and his family were moved from CA to the Rohwer, AK, Relocation Center in 1942. The family members lived in Arkansas about eight months. **Following the refusal of Takei's parents to swear loyalty to the United States, the family was moved to the maximum-security camp at Tule Lake, CA, where they stayed until 1946.**

NPR / On the Media:

- http://www.onthemedia.org/story/conversation-george-takei/
- Talks with great clarity about the FDR executive order and the loyalty oath...

Assembly Centers, Relocation Centers, Justice Department Internment Camps, & Citizen Isolation Camps



Background:
Tule Lake Segregation Center

Background:
Tule Lake Unit, National Monument

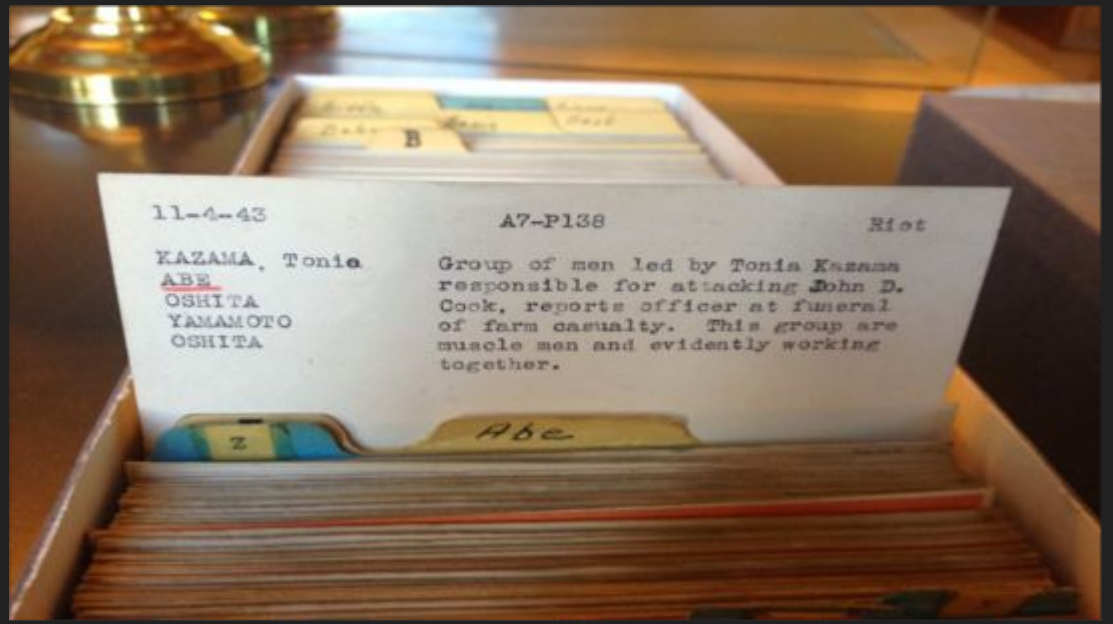
Welcome to the Tule Lake Unit of WWII Valor in the Pacific National Monument

The Tule Lake Unit, WWII Valor in the Pacific National Monument includes both the the Tule Lake Segregation Center, the largest and most controversial of the sites where Japanese Americans were incarcerated during World War II, and Camp Tulelake, which was first a Civilian Conservation Corps camp, then a prisoner of war camp, and finally an additional facility to detain Japanese Americans.

Tule Lake Segregation Center ›

Camp Tulelake ›

Incident Cards



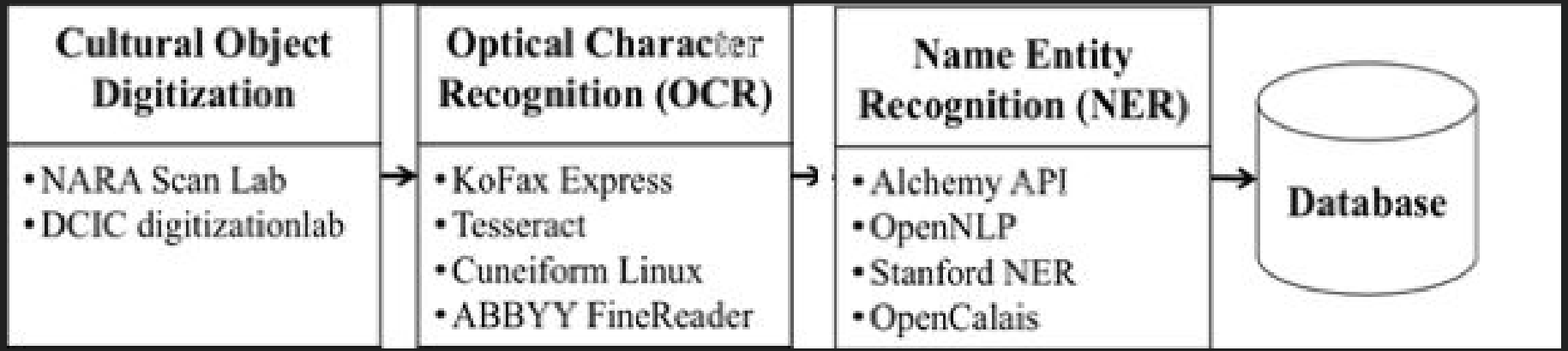
Digital Curation Workflow for Incident Cards by NARA

- Scan and OCR the "incident cards".
- Extract the names from the OCR text and capture in a database.
- Check the names against the full database of internees (also part of RG210), which includes birthdates.
- Identify names as being those of juveniles (under 18) and redact them from the database of names and from the scanned card images before their delivery to the iSchool DCIC Center.

Acknowledging the 2015 support of NARA's:

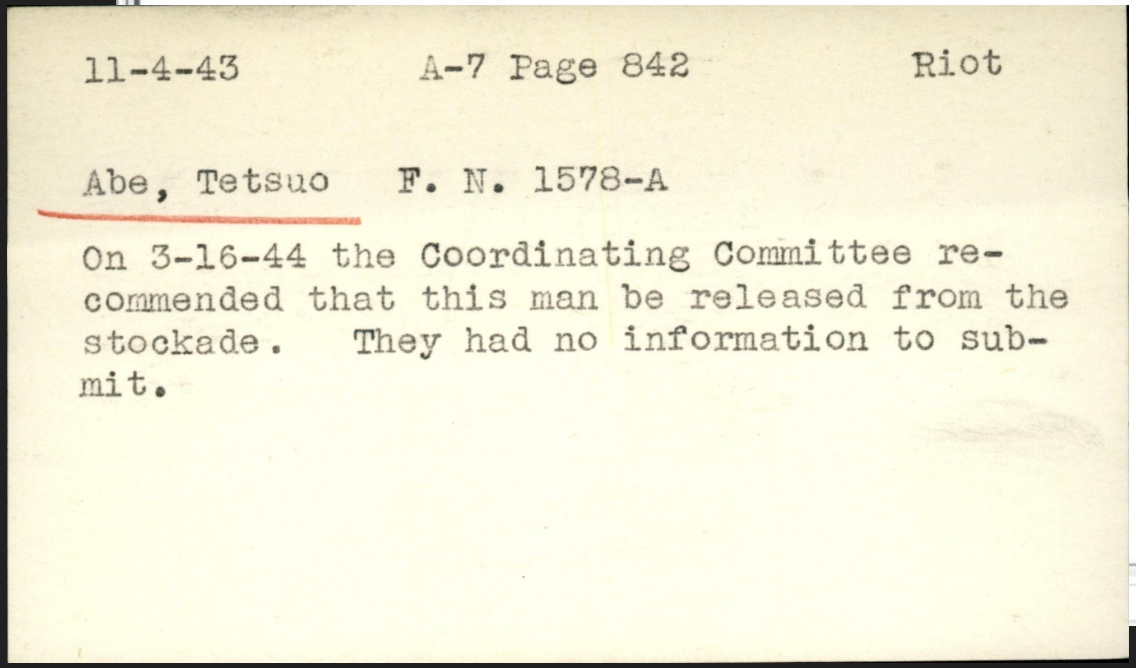
- **Office of Innovation:** Pam Wright, Markus Most, John Martinez, Denise Henderson
- **Legislative Archives & Museum Services / Exhibits Dept.:** Michelle Farnsworth
- **Federal Records Center, St. Louis:** Richard Morgan, Bob Marsh
- **Research Services:** Chris Naylor & Martha Murphy *[FOIA b(6) review]*

Digital Curation Processing Workflow



Phase II

- **Developing tools for optimized scalable input and processing** for this and other NARA collections
- **Automatic content extraction** (recognition of named entities, case ids and events) -- using GATE software
- **Structuring information** from notecards (XML Note Card document type)
- **Integration of records through analytics and visualization**

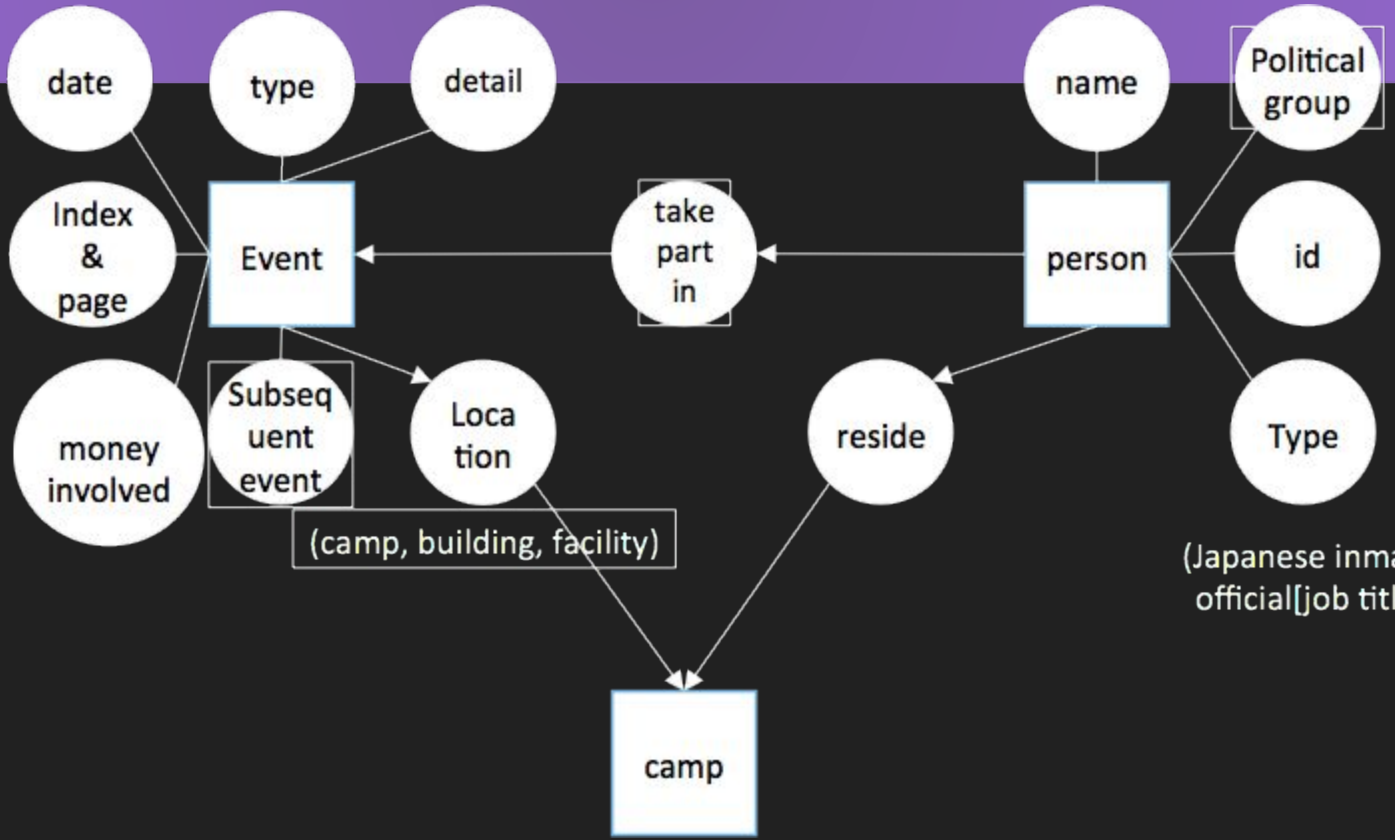


11-4-43 A-7 Page 842 Riot

Abe, Tetsuo F. N. 1578-A

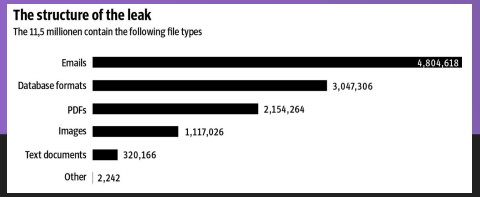
On 3-16-44 the Coordinating Committee re-
commended that this man be released from the
stockade. They had no information to sub-
mit.

Incident Card Modeling

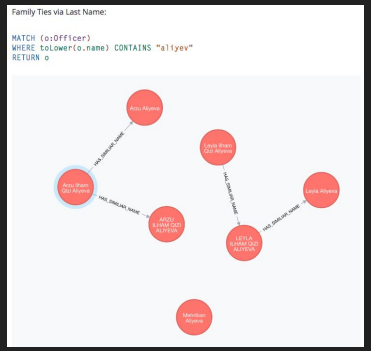


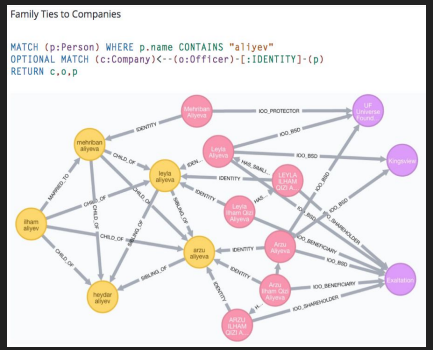
Analyzing the Panama Papers with Graph Databases:

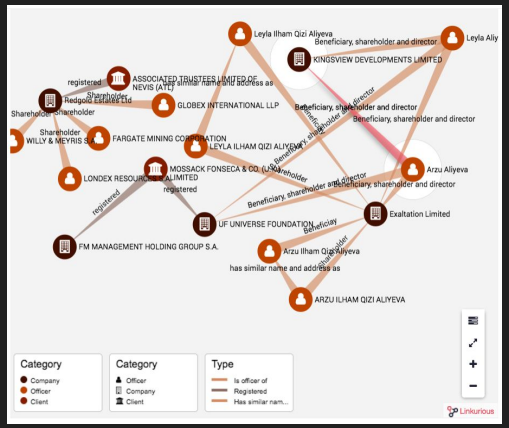
https://neo4j.com/blog/analyzing-panama-papers-neo4j/



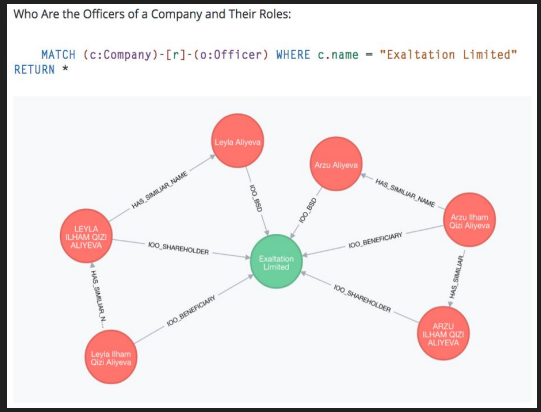
1. Acquire docs
2. Classify docs
   - Scan / OCR
   - Extract metadata
3. Data modeling
4. NER extraction
5. Relationship building
6. Graph analytics

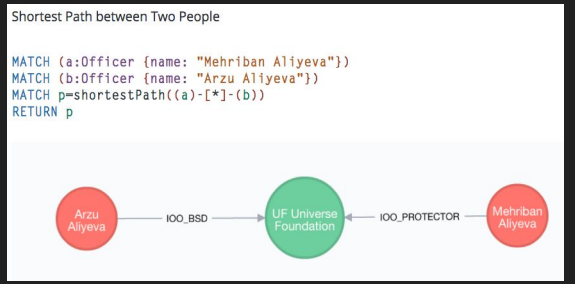






family of the **Azerbaijan's President**



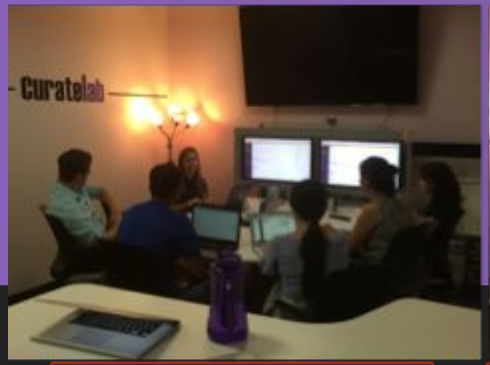


Digital Curation PROJECT OBJECTIVES:

(1) **Gain new digital skills**

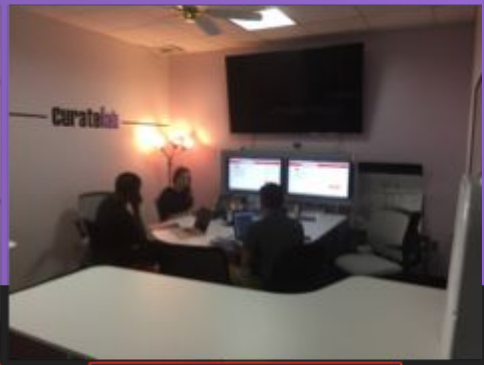
(2) **Conduct interdisciplinary research**

… at the intersection of archives, big data, & analytics



REFUGEE NARRATIVES (#8)
St. Louis Voyage

COMMUNITY DISPLACEMENT (#7)
Human Face of Big Data

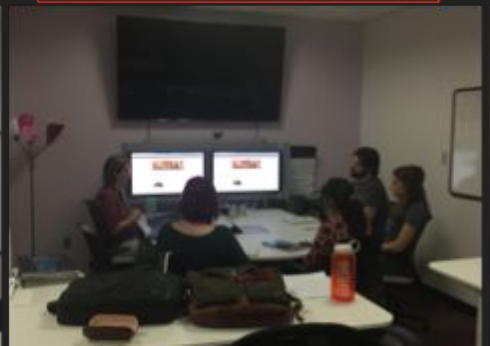


RACIAL ZONING (#11)
Mapping Inequality

DIGITIZATION LAB (#3)
Radio Log Program

CI FOR DIGITAL CURATION (#5)
Brown Dog

MOVEMENT OF PEOPLE (#5)
Overseas Pension



CITIZEN INTERNMENT (#5)
Japanese-American WWII Camps

How Each Project Is Related to Computational Archival Science (CAS) Themes:

Project	Computational Linguistics	Data Modeling & Evolutionary Prototyping	Graph Analytics	Crowdsourcing	GIS
1. Human Face of Big Data [Community Displacement]		X		X	X
2. Mapping Inequality [Racial Zoning]	X			X	X
3. St. Louis Voyage [Refugee Narratives]		X	X		X
4. World War II Japanese Camps [Citizen Internment]	X	X	X	X	X

2016 IEEE International Conference on Big Data (IEEE Big Data 2016)

http://dcicblog.umd.edu/cas/ieee_big_data_2016_cas-workshop/

Workshop Title: Computational Archival Science: digital records in the age of big data

Venue: Thursday, Dec. 8, 2016 @ Hyatt Regency Washington on Capital Hil

- **Analytics in support of archival processing**, including appraisal, arrangement and description.
- **Scalable services for archives**, including identification, preservation, metadata generation, integrity checking, normalization, reconciliation, linked data, entity extraction, anonymization and reduction.
- **New forms of archives**, including Web, social media, audiovisual archives, and blockchain.
- **Cyber-infrastructures for archive-based research** and for development and hosting of collections
- **Big data and: archival theory and practice / construction of memory / provenance**
- **Crowdsourcing** and archives
- **Specific big data technologies** (e.g. NoSQL databases) and their applications
- **Corpora and reference collections** of big archival data
- **Linked data** and archives

The DCIC is Pursuing a CAS Training / Teaching Agenda

There is a need to :

- create innovative classes that emphasize new modes of collaboration, and interdisciplinary work.
- blend elements of archival thinking and computational thinking:
  - problem solving that uses modeling, decomposition, pattern recognition, abstraction, algorithm design, and scale.
- develop inter-disciplinary iSchools with faculty from Computer Science, Archival Science, and Data Science.
- develop extensive hands-on experience working with cyberinfrastructure to carry out archival functions.

WE WELCOME PARTNERSHIPS -- CONTACT US!

NEW

DRAS-TIC

Digital Repository At Scale - That Invites Computation [To Improve Collections]

GOAL: Build out the open source DRAS-TIC platform into a horizontally scalable archives framework **serving the national library, archives, and scientific data management communities**

- Product of a 2-year startup by partners, Archival Analytics Solutions Ltd.
- Scaling to billions of files and beyond
- Interfaces:
  - Web client
  - Command-line client
  - REST storage API (CDMI) industry standard
- Key-value metadata
- Listener mechanism
- Python source on GitHub (Open AGPL license)
- Apache Cassandra database (CERN, eBay, GitHUB, Hulu, Instagram, Netflix, Twitter…)

WE WELCOME PARTNERSHIPS -- CONTACT US!

Practical Digital Curation Skills for Archivists in the 21st Century

Moving forward, build skills in the following areas:

- Digitization
  - Improving image quality
  - Transcription
- Basic Programming
  - Python / Javascript
- Data and Document Management
  - Building collections
  - Digital Storytelling
- Project Management

Questions?

http://DCIC.umd.edu

Email: dcic.ischool@gmail.com