

## Computational Archival Science (CAS) Projects:

### 1. Legacy of Slavery

In partnership with the Maryland State Archives, the project explores using digitized records from the Legacy of Slavery collections (Manumissions, Certificates of Freedom, Runaway Slave ads, Census data, etc.) in order to explore visualization, linking, and interpretation of historical records.

#### Public event:

Students participating in this project will be invited to a 2-day Data Challenge at the Maryland State Archives in Annapolis, MD. The event will take place on Monday and Tuesday, Oct. 28 / 29, 2019. The event will mirror a June datathon held at The National Archives (TNA) in Kew, UK (<https://dcicblog.umd.edu/cas/2019/07/16/uk-us-computational-archival-science-datathon-in-london/>). Students participating in this event will be participating in two ongoing funded research projects, and interacting with the research partners:

- *IRCN-CAS: International Research Collaboration Network in Computational Archival Science* (<https://computationalarchives.net/>) [UK AHRC]
- *Developing a Computational Framework for Library and Archival Education* (<https://dcicblog.umd.edu/ComputationalFrameworkForArchivalEducation/>) [US IMLS]

### 2. Japanese American WWII Camps

In partnership with **Densho.org**<sup>1</sup>, the project explores extraction and visualizing techniques to contextualize WWII camp records that have not been made available to the public before. So this is making history and experimenting with innovative digital techniques.

#### Public event:

Students participating in this project will get to showcase their work on Wednesday, Oct. 30 from 5:30 to 7:30 p.m. on campus at a public event called "Resistance at Tule Lake: A Conversation with the Filmmaker and iSchool Digital Curators" (<https://ischool.umd.edu/events-type/resistance-tule-lake-conversation-filmmaker-and-ischool-digital-curators>).

\*\*\*\*\*

#### Rules of Engagement:

- Students will have until **Fri., Aug. 23** to fill out a survey with project preference(s), including a short bio, and a statement of interest. This is open to all students.
- **Response survey link:** <https://tinyurl.com/y46jr9yn>
- There is limited seating (we will try to form up to 3 or 4 teams of 4 to 5 students each for each project). You will be notified by Monday, Aug. 26 and we will try to set up a weekly meeting time after that.
- Student commitment is to meet once a week for 8 consecutive weeks leading up to the public event, to work in teams under the guidance of Richard Marciano (DCIC Director: see <http://dcic.umd.edu>), to be able to attend the public event, and to commit to investing up to 4 or 5 hours per week during that period.

#### Benefits:

- Engagement in research projects and membership in the DCIC Center.
- Opportunity to develop innovative digital skills on real-world projects.
- Opportunity to add to your portfolio, and resume and network.
- Opportunity to be mentored by national and international professionals.
- Depending on goals and progress, opportunity to co-write a research paper.

1. <http://densho.org>: A grassroots organization dedicated to preserving, educating, and sharing the story of World War II-era incarceration of Japanese Americans in order to deepen understandings of American history and inspire action for equity.

# Fall 2019 DCIC Project Booklet

## 1. Legacy of Slavery

### Goals and Scope

The Legacy of Slavery in Maryland is a major initiative of the Maryland State Archives. The program seeks to preserve and promote the vast universe of experiences that have shaped the lives of Maryland's African American population. From the day that Mathias de Sousa and Francisco landed in St. Mary's county aboard the Ark and the Dove in 1634, Black Marylanders have made significant contributions to both the state and nation in the political, economic, agricultural, legal, and domestic arenas. Despite what often seemed insurmountable odds, Marylanders of Color have adapted, evolved, and prevailed.

### Projects from Spring 2019

1. Computing the Legacy of Slavery: Applying Computational Thinking to an Archival Dataset: See: <http://cases.umd.edu/github/cases-umd/Legacy-of-Slavery/blob/master/index.ipynb>
2. Diverse Connections: Making the 1850-1870 Calvert County Census Come Alive!

**Diverse Connections: Making the 1850-1870 Calvert County Census Come Alive!**  
By Burkely Hermann, MLIS (Expected Fall 2019)

**dcic** digital innovation center

This work has been supported by the Maryland State Archives and the Michael J. Katz Foundation through a Legacy of Slavery research internship under the auspices of the Digital Calverton Preservation Center.

**Census Surnames in Calvert County (1850-1870)**  
Surnames of Black peoples (1850-1870) within Calvert County's District 2

**Surnames of Multiracial people (1850-1870) living in Calvert County's District 3**

The surnames, marks, and writings on U.S. Federal Censuses, from 1850-1870 not only have meaning for people today, when the family historians or researchers, but were meaningful to those living during that period, especially for underrepresented people. This poster aims to make those census documents come to life through a series of visualizations I created this semester, as part of a research cohort at the DCIC, in concert with those wonderful fellow students from the University of Maryland (Jane Sliwa, Ali Ghaffar, and Chrissy Peery) with a faculty advisor of Prof. Farhan and assistance from Noah DeBart of the DCIC.

• Use the records of Black and Mixed Race Marylanders gathered by the Maryland State Archives (MSA), for inclusion in their Legacy of Slavery project, and visualize the results.

During the past spring 2019 semester, I have:

- created 25 visualizations from Calvert County Census data from 1850-1870, 12 of which are shared on this poster;
- used tools such as Infolingo to create a word cloud of surnames
- used PivotTables to create bar charts of people to show gender distribution
- draw connections between census records and land records
- used Datawrapper to create charts showing the age, race, and occupation of Black and Mixed-Race Marylanders
- used Microsoft Excel and ChartGo to chart real and personal estate value

Other than some inaccuracies in the data, like an incomplete data set for Calvert County District 2, I did not visualize the hundreds of Black individuals whom the census later listed with no occupation because that the categorization may have been locally imbedded. This poster also does not display those whom had specific claimed mental deficiencies because it is similar in format to the other Excel charts or the two possible stories that involve connections with census, land, and genealogical records due to their limited scope.

With that, I hope you enjoy this poster, which is just the beginning of visualizing this data and creating diverse connections!

**Visualizing demographics of Black and Mixed Race Marylanders (1850-1870)**  
Average age and racial classification for Black and Mixed Race Marylanders within Calvert County District 3

Year	Age	Racial Classification
1850	21.7	21
1860	22.2	22
1870	22.1	22

**Average age and racial classification for Black Marylanders within Calvert County District 2**

Year	Age	Racial Classification
1850	21.9	21
1860	22.0	22
1870	22.1	22

**Jobs of Mixed Race and Black Marylanders (1850-1870)**  
Occupations of Black Marylanders in Calvert County District 1

Year	Occupation	Count
1850	Unemployed	10
1850	Domestic	5
1850	Farmer	2
1850	Other	1
1860	Unemployed	12
1860	Domestic	8
1860	Farmer	3
1860	Other	2
1870	Unemployed	15
1870	Domestic	10
1870	Farmer	4
1870	Other	3

Occupations of Mixed Race Marylanders in Calvert County District 3

Year	Occupation	Count
1850	Unemployed	8
1850	Domestic	4
1850	Farmer	2
1850	Other	1
1860	Unemployed	10
1860	Domestic	6
1860	Farmer	3
1860	Other	2
1870	Unemployed	12
1870	Domestic	8
1870	Farmer	4
1870	Other	3

**Lessons learned**  
Looking back, there are some parts of this process I wish I had done differently:

- Rather than handwriting, with the help of both the Microsoft Excel, of surnames, occupations, and race, it would have been more efficient to use OpenRefine
- Using Infolingo to represent people, one of their fun features, as would be nice reflection from trying to painstakingly draw the people myself, but I did not realize this feature was available until after I had completed the visualization in Paint.NET

At the same time, I wish:

- only partially satisfied with the charts created about occupations, race, and average age, but I'm not sure how wise to visualize the data at this time
- a bit disappointed at how my visualizations connecting land and census records turned out, as I was unable to figure out how to add layers to an image using PhotoGrip, so I stuck with Paint.NET

Furthermore, I wish there was a way to connect to existing maps, apart from the 1870 map of Calvert County which displays districts, but I was unable to find any maps from 1850 and 1860. While I realize these shortcomings, the work I completed this semester is vitally important, allowing the census data to fully come alive rather than be static.

MARYLAND'S ISCHOOL

## Fall 2019 DCIC Project Booklet

### 3. Legacy of Slavery: Constructing Free Black Marylanders' Lived Experience through Government Documents:

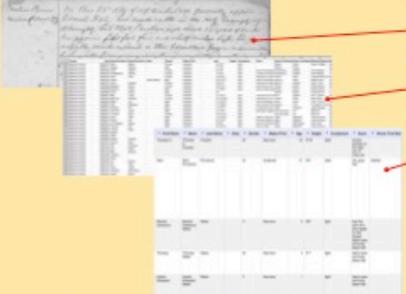
# Legacy of Slavery: Constructing Free Black Marylanders' Lived Experience through Government Documents

## Benjamin Shaw

Certificates of freedom were documents that free-born and emancipated black Americans in slave states had to carry with them at all times, to prove to law enforcement of a moment's notice that they were legally free people. As a slave state, Maryland issued these documents and enforced this practice on free Black Marylanders, with the constant underlying threat of re-enslavement. The Maryland State Archives provided the Digital Curator Innovation Center with a ledger of these records, taken down at the Baltimore County Courthouse in the 1840s.

Documentation from black Marylanders' perspectives, especially working-class black Marylanders, is rare from this time period. For most, we have very few records of their lives at all. This is why these certificates of freedom, almost two centuries later, have become so important. The level of personal detail in these records means that we can glean tremendous detail from these dehumanizing documents. Using OpenRefine and Palisade, I was able to provide a (very brief) biographical sketch of each person described by these records. The hope here is to move towards a person-first depiction of black lives lived in slavery-era Maryland, and to use these documents to begin to generate narratives from datasets of details.



The project began with transcription of the handwritten ledger of certificates of freedom. Clerks would write two copies of the certificates, giving one to the person it described and writing a second in the court's ledger, for protection against forgeries.

Specific, standardized pieces of information were transcribed into an Excel spreadsheet, to allow for easy work with basic building blocks of the data. The rest was added to a "Notes" section; this unstructured data ended up providing a large portion of the information in the narratives and graphs on this poster.

Once this data was compiled, I branched it in OpenRefine to allow for data cleaning, consolidation of different data points, automated formatting of dates, and automated generation of the captions to the right.

Using OpenRefine allowed me to work with OREL (General Refine Expression Language) to auto-generate certain portions of a narrative template. The intent to the right shows the formula I used for these narratives. The gallery of narratives was generated using Palisade.



[Name: "Yakov"] was born [Month: "March"] [Status: "male"] in [County: "County of St. Mary"] [Age: "18"] [Sex: "male"], [Height: "5'6"] [Age: "18"], and was living near Baltimore, MD. [Certificate Date: "1841"] [Personal Pronoun: "she"] came to the Baltimore County Courthouse [Age: "18"] [Sex: "male"], [Height: "5'6"] [Personal Pronoun: "she"].

The problem of "distinguishing marks" in addition to the type of information that we state on government IDs today - name, gender, height, age - is the white clerk writing the certificate would mark down anything about the bearer that he considered to be distinguishing. Other than was a scar, a mole, or a personal quirk. The goal, even more than identifying an individual, was to describe the bearer so specifically that they could not share the certificate with anyone else. Five people sharing documentation with enslaved Marylanders was a prevalent form of resistance, and the state government of Maryland was eager to stop it from happening.

This section in the form was meant as a tool to curtail the spread of freedom in Maryland. However, this field often gives us the small, personal details that are the most evocative parts of these documents, and they're some of the best chances we have of glimpsing the lives of the people who held these certificates. They give hints to the life experiences a person might have had, in terms of the trade they knew, the life they led, or the labor they were forced to perform. Some examples:

- Thomas A. Franklin had an scar/scarred on his right arm.
- Christiana Sudder had scars on her right hand and elbow.
- Henry Johnson wore rings in his ears.
- Samuel Lambrey had a crooked finger on his left hand.

These would seem to be ideal for making to create the stories above. However, this field is unstructured text, which makes it very complicated to work with. Furthermore, the language the clerk used to describe the applicant was often extraordinarily dehumanizing. These would be difficult to read out-into a narrative that does not perpetuate the dehumanization of the people whose stories we are trying to resurrect. To make this into a useful part of these stories would require the development of a controlled vocabulary, some in-depth data cleaning, and thoughtful work on what types of information to pull from the text field.



The map to the left was also generated by uploading the certificates of freedom data to Palisade. It shows the concentration of home counties of the subjects of the certificates. Most were from Baltimore or Baltimore County, as that was where the courthouse resided. However, in the four-year span that the dataset covered (1841-1845), people born in almost every county in Maryland came to the Baltimore County Courthouse to obtain certificates of freedom. Whether having moved voluntarily as a free person, or having been forcibly displaced under slavery, people from all over the state ended up beginning their lives as free Marylanders in Baltimore. (This line at the very bottom represents two certificate-holders who were born in South Carolina, the only people in the dataset who were not native Marylanders.)

The map was generated by extracting the names of certificate-holders' home counties from the unstructured "Notes" field in the original spreadsheet. The county names were then converted into rough coordinates, and fed into Palisade's point-to-point function.

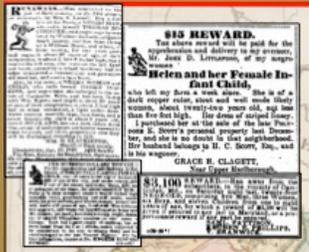
The work has been supported by the Maryland State Archives and the Michael J. Kurzi Foundation through a Legacy of Slavery research internship under the auspices of the Digital Curation Innovation Center. Special thanks to Kathia Fankon and Nash O'Brien for their constant support and guidance through this project.

### 4. Families in Flight: Runaway Slaves and their Children 1783-1859:

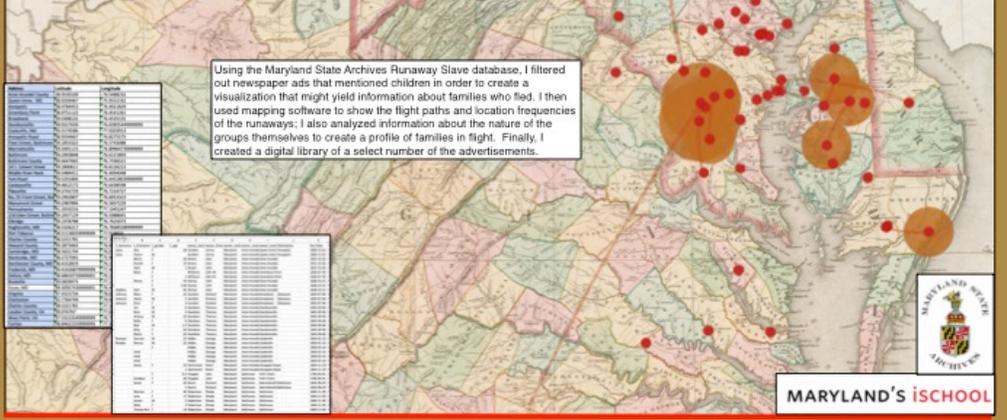
# Families in Flight: Runaway Slaves and their Children 1783-1859

## Christiana P. Perry, MLIS '20 May 8, 2019



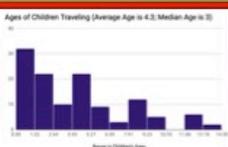


Using the Maryland State Archives Runaway Slave database, I filtered out newspaper ads that mentioned children in order to create a visualization that might yield information about families who fled. I then used mapping software to show the flight paths and location frequencies of the runaways; I also analyzed information about the nature of the groups themselves to create a profile of families in flight. Finally, I created a digital library of a select number of the advertisements.





Bar of Families Running



Ages of Children Traveling (Average Age is 4.3; Median Age is 3)



Number of Children by Age

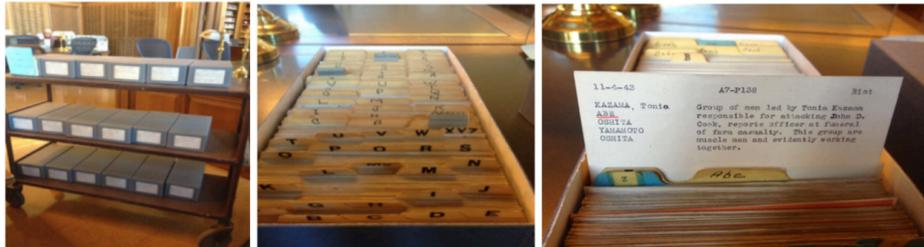
The work has been supported by the Maryland State Archives and the Michael J. Kurzi Foundation through a Legacy of Slavery research internship under the auspices of the Digital Curation Innovation Center.

## 2. Japanese American WWII Camps

### Goals and Scope

During World War II, over 120,000 Japanese Americans were relocated and jailed across 10 camps. The National Archives maintains record series related to the *War Relocation Authority (WAR)* agency that oversaw the incarceration. Among these are “Internal Security Case Reports” prepared by Relocation Center staff relating to alleged cases of disorderly conduct, assault, theft, loss of property, and accidents. This project focuses on the “Internal Security Cases” index cards. Each card includes a case number, type of charge, names and addresses of persons involved, time and place where the incident occurred, and account of the incident, and refers to a more detailed case file (for which access is restricted). See:

- <http://ddr.densho.org/names/>
- <https://catalog.archives.gov/id/1264228>
- <http://local.ads.nwsources.com/ads/FlippingBook/2015/Q2/Densho/html/>



### Projects from Spring 2019

1. Revisiting Segregation through Computational Treatments: the Case of the WWII Japanese American Tule Lake Segregation Center. See: <http://cases.umd.edu/github/cases-umd/Japanese-American-WWII/blob/master/index.ipynb>
2. Mapping Japanese-American WWII Incarceration Camp Records:

## Mapping Japanese-American WWII Incarceration Camp Records

INST 747 Dr. Richard Marciano and Dr. Ken Heger

Ryan Tetter [MLIS]  
Connor Mullane [MLIS]

Tule Lake Camp Map

With detailed information on the names and addresses referenced in the Tule Lake incident cards, we are able to visualize the data by geolocating the persons involved.

Normalized Incident Cards with Names Redacted

Through the use of OpenRefine and the VIB-Bits plugin, we cleaned the data from the Tule Lake incident cards and created a column of barracks IDs that matched our data in QGIS. The names have been redacted due to privacy concerns.

All Boxes Georeferenced onto Camp Map

Each dot on our georeferenced map represents one or more individual incident card that is linked to that location. Because each barrack could house four families, there are often multiple cards referenced in each dot.

**Computational Thinking Practices**

**Data Practices**

- Manipulating Data
- Analyzing Data
- Visualizing Data

**Modeling & Simulation Practices**

- Designing and Constructing Computational Models

**Computational Problem Solving Practices**

- Developing Modular Computational Solutions
- Troubleshooting and Debugging

**Tule Lake Incident Cards**

**Figure 1**

Fig. 1  
These lines of code show how the residence IDs were normalized to match those in QGIS. This was a simple splice function that required some correction by hand, but still resulted in the majority of our data being set up to be quickly fed into QGIS.

**Figure 2**

Fig. 2  
Coordinate data imported using the VIB-Bits cross-sheet plugin, allowing our incident card data to be displayed directly in QGIS.

**Further Research**

We have additional plans to expand our computational model in QGIS:

- Include links to webpages and historical documents to make the map more interactive
- Simulate landscapes in 3D
- Manually add labels
- Display multiple canvases
- Run Python 3 scripts to enhance functionality beyond traditional commands

MARYLAND'S ISCHOOL

## Fall 2019 DCIC Project Booklet

### 3. Visualizing Tule Lake: A Data-Driven Approach to Understanding Japanese American Internment During World War II:

# Visualizing Tule Lake: A Data-Driven Approach to Understanding Japanese American Internment During World War II

Tracee Haupt & Mohammad Hanaee INST747 Dr. Richard Marciano & Dr. Kenneth Heger

Tule Lake was one of ten "relocation camps" built to imprison Japanese Americans forcibly removed from the West Coast during World War II. Of the nearly 120,000 Japanese Americans imprisoned after Executive Order 9066 was issued in early 1942, about 30,000 were held for a time at Tule Lake. The population of Tule Lake was in flux during the four years (1942-1946) it was open, as people were transferred from one camp to another, but a peak population of nearly 19,000 people made Tule Lake the largest of the ten camps. Our project analyzed data from the Final Accountability Roster (FAR) to find out more about who was held in the camp. The FAR was transformed from a print record to a spreadsheet, and the data was cleaned to make it more accurate and analyzable. Data could never tell the full story of life in the camp, but it can be a valuable supplement to the historical record.

#### Years Spent Incarcerated at Tule Lake

#### Gender of the Incarcerated

#### State of Origin of the Incarcerated

Most were from California, but some were from Oregon, Washington, and Hawaii.

#### Citizenship Status

Approximately 69% of the incarcerated at Tule Lake were American citizens, while 31% were considered "aliens," or non-citizens. Of those who had citizenship, 27% at Tule Lake renounced their citizenship. There were reasons for rejecting U.S. citizenship that did not involve disloyalty—some did so in protest of the camps, others were afraid of being separated from their families if non-citizens were deported or of being released into hostile white communities while the war was still being waged, and some were under duress from pro-Japan buffies in the camp.

#### Deaths in the Camp: 314

One of the deaths at Tule Lake was the unsolved murder of Co-Op General Manager, Yuzo Hironi. An avenue for future research may be learning more about the lives of others who died at the camp.

#### Age of the Incarcerated

9127, or 38.1% of the population of Tule Lake was under the age of 18 when they first entered the camp.

#### Births in the Camp: 899

One of the babies born in the Tule Lake camp was Satsuki Ina, who became a community activist, writer, and filmmaker. She has made two documentaries about Japanese American internment, including "Children of the Camps." (Photo credit: satsukina.com)

### 4. The Power of Controlled Vocabulary: Drawing Narratives of Internment from Big Data:

# The Power of Controlled Vocabulary: Drawing Narratives of Internment from Big Data

Margaret Rose Hunt and Margaret McCreedy

#### Historical Context

After the attack on Pearl Harbor in 1941, President Franklin D. Roosevelt signed Executive Order 9066 in February 1942. As a result, over 120,000 people were forcibly detained and sent to internment camps, leaving behind jobs, businesses, and homes.

Designated as a segregation zone for "enemy aliens," the Tule Lake internment camp at once housed 10,000 people, specifically designed to contain the most "dangerous" and "disloyal" enemy alien. Japanese Americans lived in a desolate compound under martial law until 1946.

Under martial law, anything from taking water bottles, a car accident, or peaceful protests carried out by Japanese citizen groups like the Hoshi-Dan and the Seinen-Dan, were considered crimes, and each "incident" was recorded in a coded format on an index card. The data derived from these incident cards documents the experiences of individuals incarcerated at Tule Lake, so each "location" in this data refers to an event that a person lived through.

#### Context of Our Project

These incident cards were digitized, and the data was made available to a group of students in INST 747. Researchers at Maryland Digital Curation, for several projects that exist, bring new historical context to the incident cards. Data was pulled from incident cards using OCR and NER, and sorted into offense categories based on where the information was on the card. One of these categories was the "offense" category, which described either a crime committed by an administrative action. However, the data had numerous spelling mistakes, incomprehensible abbreviations, and inconsistent capitalization. To be usable in an automated and computational way, data cleaning and harmonizing needed to occur.

#### Incomplete Offense Information

Even with standardized offense types created through the controlled vocabulary, several different categories were used to document the same types of offenses. One of the most prevalent was the categorization of political dissent. Three possible separate categories emerged: Infraction of Project Regulations, Political Activity in Colony, and Suspicious Activity.

The categories in the left image were pulled after data cleaning, when "Violation of Project Regulations" was changed to "Infraction of Project Regulations" because that was more commonly used.

#### Creating the Controlled Vocabulary

Using OpenRefine to cluster the offense types, we noticed significant variability in the way camp administrators wrote out the same offenses. The image below shows all of the different ways a transfer to North Dakota were written; this was cleaned to one standard term using OpenRefine's cluster tool.

#### Using the Controlled Vocabulary

After harmonizing the terminology used to describe transfers, we noted that many of the transfers occurred on several specific dates. We discovered one of the last major transfers of internees to Bismarck, ND was of several hundred "reservants" being prepared for deportation to Japan. Above is an article from Tule Lake's newspaper, *The Newell Star*.

*The Newell Star*, 2-15-46

#### Finding the Details

While the details in the "offense" column are lacking, other columns give more information that can help organize the data into more descriptive offense types. These "subtypes" could be reclassified as political activity or under the "offense" group for the reports about the Hoshi-Dan or the Seinen-Dan based on this additional information.

With this additional context provided by the *Newell Star*, the true meaning of the term "Violation of Project Regulations" is revealed. The frequency with which this offense type appears in the data indicates consistent political activity and resistance on behalf of a large portion of the Tule Lake population.