




Using AI/Machine Learning to Extract Data from Japanese American Confinement Records

Mary Elings, University of California Berkeley, Berkeley, CA, USA, melings@berkeley.edu 
Marissa Friedman, University of California Berkeley, Berkeley, CA, USA
Vijay Singh, Doxie.AI, San Jose, CA, USA

Abstract

Purpose: This paper examines the use of Artificial Intelligence/Machine Learning to extract a more comprehensive data set from a structured “standardized” form used to document Japanese American incarcerated during World War II.

Setting/Participants/Resources: The Bancroft Library partnered with Densho, a community memory organization, and Doxie.AI to complete this work.

Brief Description: The project digitized the complete set of Form WRA-26 “individual records” for more than 110,000 Japanese Americans incarcerated in War Relocation Authority camps during WWII. The library utilized AI/machine learning to automate text extraction from over 220,000 images of a structured “standardized” form; our goal was to improve upon and collect information not previously recorded in the Japanese American Internee Data file held by the National Archives and Records Administration. The project team worked with technical, academic, legal, and community partners to address ethical and logistical issues raised by the data extraction process, and to assess appropriate access options for the dataset(s) and digitized records.

Received: November 15, 2023 **Accepted:** February 5, 2024 **Published:** March 6, 2024

Keywords: libraries, artificial intelligence, AI, machine learning, archives, Japanese American incarceration, World War II

Citation: Elings, Mary, Marissa Friedman, and Vijay Singh. “Using AI/Machine Learning to Extract Data from Japanese American Confinement Records.” *Journal of eScience Librarianship* 13 (1): e850. <https://doi.org/10.7191/jeslib.850>.

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, and new creations are licensed under the identical terms.

See <https://creativecommons.org/licenses/by-nc-sa/4.0>.

∞ OPEN ACCESS

Abstract continued

Results/Outcome: Using AI/machine learning increased the quality of the data extracted from the digitized WWII era forms.

Evaluation Method: A comparison of the earlier dataset extracted from the 1940s's computer punch cards to the current data set extracted using AI/machine learning, the use of AI/machine learning showed marked improvement.

Project description

With funding from a 2019 National Park Service Japanese American Confinement Sites grant, The Bancroft Library digitized the complete set of Form WRA-26 “individual records” for more than 110,000 Japanese Americans incarcerated in War Relocation Authority camps during WWII. The library partnered with Doxie.AI to utilize AI/machine learning to automate text extraction from over 220,000 images; our goal was to improve upon and collect information not previously recorded in the Japanese American Internee Data file held by the National Archives and Records Administration (NARA). The project team worked with technical, academic, legal, and community partners to address ethical and logistical issues raised by the data extraction process, and to assess appropriate access options for the dataset(s) and digitized records.

Overview

This project offered our library the first opportunity to use AI/machine learning to improve data extraction from a digitized historical resource. Our goal was to enhance access to the information held within that resource and ultimately support emerging scholarship and computational analysis. Because the expertise did not exist in our library, we partnered with a team of data scientists. Their role was to develop a custom machine learning pipeline for the data extraction. Our role was to facilitate that work, provide guidance and content expertise to the data scientists, and review/quality control (QC) the results.

Narrative

Summary

With funding from a 2019 National Park Service Japanese American Confinement Sites grant, The Bancroft Library digitized the complete set of Form WRA-26 “individual records” for more than 110,000 Japanese Americans incarcerated in War Relocation Authority camps during WWII. The library partnered with the data scientists at Doxie.AI to utilize AI/machine learning to automate text extraction from over 220,000 images; our goal was to improve upon and collect information not previously recorded in the *Japanese American Internee Data file* held by the National Archives and Records Administration (NARA) (National Archives 2024). The project team worked with technical, academic, legal, and community partners to

address ethical and logistical issues raised by the data extraction process, and to assess appropriate access options for the dataset(s) and digitized records.

Project details

In 2019, The Bancroft Library at the University of California, Berkeley, received funding through the National Park Service Japanese American Confinement Sites grant program to digitize materials from the *Japanese American Evacuation and Resettlement Records* (BANC MSS 67/14 c) collection, specifically the complete set of Form WRA-26 “individual records” for more than 110,000 Japanese Americans incarcerated in War Relocation Authority (WRA) camps during World War II (Bancroft 2019). The five main goals of this project included:

1. Digitizing and creating a preservation copy of the Form WRA-26 records for future generations, as these records are of enduring and significant historical value;
2. Reaching consensus among community representatives and stakeholders as to how best to provide access to the Form 26 material;
3. Providing a new, more complete dataset relating to Japanese American WWII incarcerated which improves upon the errors, gaps, and omissions in the existing data file which was generated from computer punch cards created during WWII;
4. Testing, creating, and implementing workflows and tools which can help the Bancroft Library transform its growing digital archival collections into data that can be made available for computational analysis and enhanced access;
5. Building, testing, and implementing a sustainable model for integrating community input into our work in alignment with our own Responsible Access Workflows (UC Regents 2021).

The project was led by Principal Investigator (PI) Mary Elings, Interim Deputy Director and Head of Technical Services, and managed by Digital Project Archivist Marissa Friedman of the Bancroft Library. The library contracted with Backstage Library Works to digitize the original forms and with Doxie.AI to implement AI/machine learning to extract data from the digitized forms. The PI and Digital Project Archivist worked closely with Densho, a community memory organization dedicated to preserving the history of the incarceration, and the UC Berkeley Office of Scholarly Communication Services, to organize a Community Advisory Group meeting (Densho 2024). The goal of this meeting was to bring together community experts to advise on how to ethically and responsibly expand access to these sensitive records.

To understand why these collections were selected as the basis for this project, it is helpful to understand a bit of the historical context and lifespan of both the analog records and the data these records contain. From 1942 to April 1943, a census-type two-page Form WRA-26 (or “Individual Record”) was used to collect a wide range of demographic, educational, occupational, and biographical data about every Japanese American incarcerated in WRA camps during the war (Figure 1).

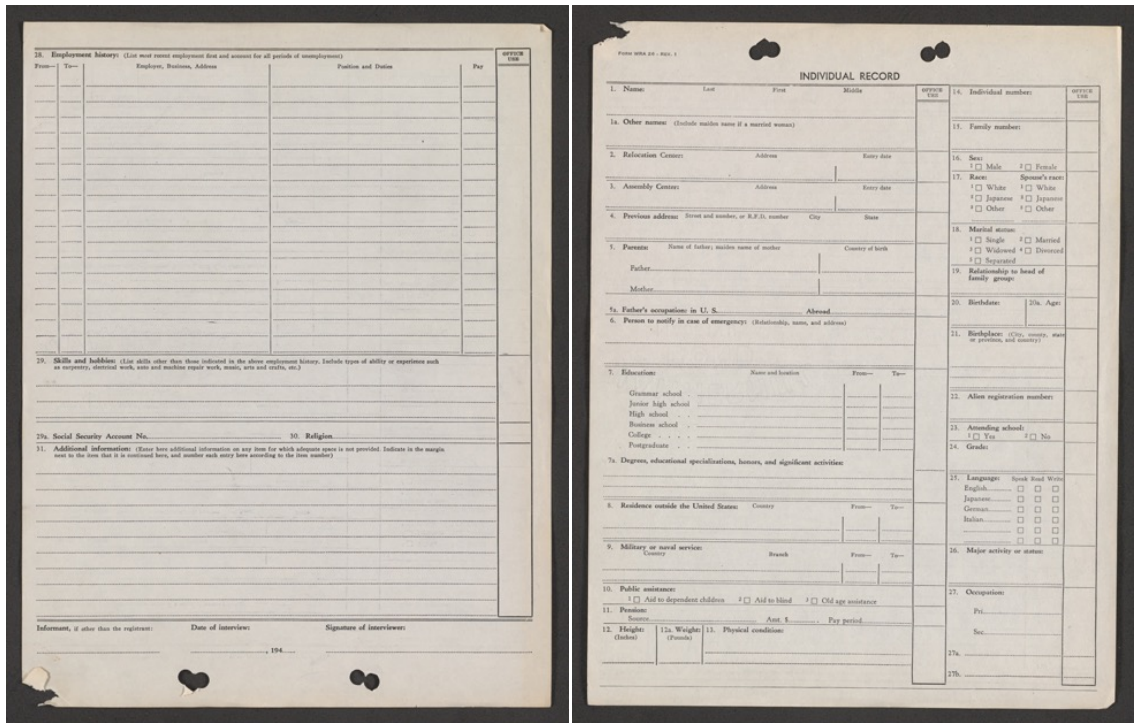


Figure 1: Blank Form WRA-26 "Individual Records" (front and back). There were several variations of this form under the same name with slight changes to location and number of fields. Courtesy The Bancroft Library.

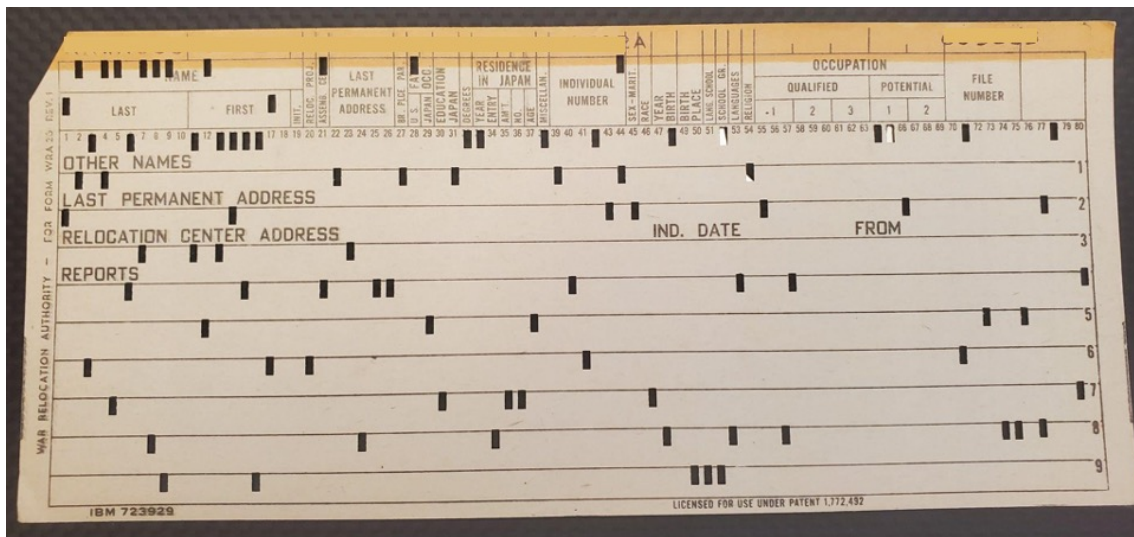


Figure 2: War Relocation Authority Computer Punch Card. Courtesy Densho.

The information included some potentially sensitive personal data and was taken under duress and without consent from forcibly relocated and incarcerated individuals. During the war, data from the Form WRA-26 records was coded by incarcerated Japanese Americans and other WRA office staffers to early computer punch cards so that the information, some of it generalized into broad categories, could be processed by tabulating machines (Figure 2).

At the conclusion of the war, a copy of the Form WRA-26 punch cards and the original type- or handwritten forms from which the punch cards were coded were deposited at The Bancroft Library along with many other WRA records. In the 1960s, the Form WRA-26 data on the computer punch cards was transferred onto magnetic tape by the library with help from the nascent UC Berkeley computer science department. The Office of Redress Administration (ORA) acquired a copy of the data from The Bancroft Library in 1988 to aid in disbursing reparations to former Japanese American incarcerated. Upon completion of the agency's work, the modified file was transferred to the National Archives. NARA published the data file it acquired from the ORA, along with extensive documentation, as part of its Access to Archival Databases (AAD) project in 2003. Referred to as the *Japanese American Internee Data File*, this datafile currently serves as an authoritative resource for genealogical information for former inmates and their family members, as well as statistical information about the incarcerated population as a whole.

Thanks to Densho, we were made aware that the more than 110,000 Form WRA-26 records held at the Bancroft were possibly the only remaining complete set, organized by camp, in existence. While digitization for preservation then became the immediate concern for these records of unique and enduring research value, in the spirit of the Archives and Collections as Data movements, the library also wanted to explore how the information in these records might be made available for computational research. Digitizing the entire corpus of Form WRA-26 records provided an opportunity to extract data from these records and create a new dataset which might rectify the gaps, omissions, and errors that are present in the existing datafile at NARA due to how the data was originally created. In order to accomplish this, we needed to transcribe and extract an enormous quantity of data from over 220,000 images, an undertaking which was not feasible given current staffing, expertise, and resource levels.

We recognized that we would need to add team members to the project with the technical expertise and experience to help us use machine learning to pull data from these records. Library staff first met the team members of Doxie.AI while they were graduate students in the Master's of Information and Data Science program at UC Berkeley's School of Information. The team of data scientists had done a previous project, BugTrAP: (Bug Transcription and Annotation Pipeline), that extracted data from images of labels on entomological specimens (Doolittle, et al. 2020). This work demonstrated their interest and experience in developing customized machine learning pipelines to transcribe text from images. Because we lacked in-house expertise, we discussed the project and ethical concerns posed by working with these records with the team of data scientists and their faculty advisor. Satisfied that this approach would yield the results we

hoped for (efficient and high quality text transcription from complex resources), the library partnered with these UCB-affiliated data scientists to automate the text extraction process using AI and machine learning. From an ethical perspective, we worked closely with the team to secure the data and address any ethical or sensitivity concerns in the data. Our collaboration with them spanned nearly two years, during which they transitioned from graduate students in UC Berkeley's MIDS program to an external vendor following their incorporation as Doxie.AI (Doxie.AI 2024). Working with the Doxie.AI team, we learned a great deal about what was possible, and impossible, is using AI/machine learning for improved text extraction, as compared to OCR or hand transcription.

At the initial stage of the project, we also consulted with experienced data publication colleagues in the Library to determine available and appropriate methods for data transfer, storage, and project management tracking. Given the sensitivity of the data, we used Box to more securely transfer digital files from the library to Doxie.AI team members, and created a private Github repository to securely store the extracted data—in CSV and JSON formats—and documentation. With a basic workflow for the transcription process in place, Doxie.AI began testing and developing a customized machine learning model for the materials which could be iteratively improved upon as we progressed through the records, moving from camp to camp. All models were supervised; reinforcement learning (RL) or unsupervised learning was not used for this project. Doxie employed model fine-tuning to achieve the best results possible given the data drift introduced by variances in the type, spacing, and placement of data from camp to camp. This variance was monitored as the project progressed, and additional data was used to fine-tune the models for a more robust performance. This process involved quite a bit of knowledge sharing between Doxie.AI and Bancroft staff, as the archivist managing the project noted key characteristics of the records and relayed what content we hoped to transcribe; meanwhile, Doxie.AI defined the parameters of what was technologically possible and continually worked to expand the pipeline's capacity and accuracy to match new observations discovered about the form and content of the records. Due to variability in structure, content, and other unique characteristics of the physical records within and across camps, the team adopted an iterative approach which handled sets of Form WRA-26 records one camp at a time. After a record set from one WRA camp was run through the pipeline, results were reviewed by Bancroft staff, and Doxie.AI attempted to integrate any feedback into future work.

A number of challenges arose throughout the process. For example, the individual records organized by camp turned out to contain a significant number of forms with entirely handwritten responses as well as a numerous forms with handwritten notations, annotations, and corrections, all of which were not able to be fully or accurately captured in the machine learning transcription process due to limitations with the technology when applied to handwritten materials. Additionally, six or seven versions of the form were used, with discrepancies in data types, content, and spacing of text on the page disrupted the ability of the ML pipeline to accurately transcribe content. Many documents also included stamps, handwritten corrections, strikethroughs, notes, and other marginalia, which presented visual noise that was difficult

for automated transcription models to handle. There are limitations to even a finely tuned and customized machine learning pipeline when applied to the complexity and inconsistency of even fielded, form-based data in archival documents.

Beyond the challenges of implementing AI/machine learning technology, the project presented a number of ethical issues due to the potentially sensitive nature of the records being digitized, records which contain PII and other information about vulnerable individuals. The project team worked with library and community partners to address ethical and logistical issues raised by the data extraction process, and to assess appropriate access options for the dataset(s) and digitized records. A critical component of this work was forming a community advisory group which met virtually in August 2022; the goal of this meeting was to garner meaningful community feedback on the work accomplished so far, and work towards some sort of consensus on whether and how to provide access to the digitized forms and dataset.

Background

This project presented an interesting opportunity to apply AI and machine learning tools to our digital collections. First, the highly structured and (we thought) consistent format of the materials being digitized lent themselves to applying an automated transcription approach. The form-based records included primarily structured and typewritten data; while transcription of historical handwritten texts is notoriously harder to achieve with the current set of tools available, machine learning tools for transcribing typewritten text are fairly advanced and have shown demonstrably good results. Additionally, machine learning offered a reasonable solution to logistical constraints; the sheer size of the corpus of digitized materials in question (over 220,000 files) and current staffing and resource levels did not practically accommodate labor-intensive and costly manual transcription. Adopting automated tools appeared to be a more efficient and less staff- and cost-intensive approach for achieving a mass transcription project in a timely manner as compared to traditional hand transcription services or direct optical character recognition. We also considered public crowdsourcing tools and digital platforms such as Zooniverse and From the Page, but decided against these options due to a few reasons. First, the sensitive nature of the records potentially precluded mass public access, and second, the library was not at that time in a position to organize and manage a public crowdsourcing transcription project. The library looked to AI to help provide a significant lift in transcribing data from the digitized records to create a dataset which would improve upon and expand the existing Japanese American Internee Datafile held by NARA.

As mentioned above, the existing data file available at NARA, based on the 1940s punch cards, demonstrates a number of key limitations. The number of migrations the data went through over time introduced or compounded errors and inaccuracies from the original data collection process. Even more importantly, a significant amount of detailed information collected in the original paper forms is missing from or was generalized in broad categories in the existing data file. In some instances, entire fields were absent from the NARA dataset, including significant activities, skills, hobbies, educational and employment history, and

the field for additional information. In other cases, the granularity of information was generalized through the act of coding responses to a pre-set number of categories or datapoints; for example, occupations were coded to a prescribed set of classifications, producing a loss of significant detail supplied in the original forms. We hoped that AI/machine learning could help us to recover as much of this information as possible, while transforming it into formats more readily accessible for computational research.

Inspiration for this project came from many sources. UC Berkeley's Digital Humanities program launched efforts in 2012, supporting computational "research ready" access to archival materials in a number of early pilot efforts that led the way to this project (Berkeley Center 2024). The concept of "research ready" data is when digitized corpora are machine readable, queryable, maintains original structure, annotated, and linked to other data on the same topic (Adams 2017). The principles and documentation produced by participating in the *Always Already Computational: Collections as Data* grant-funded initiative (2016-2018), which has been succeeded by the *Collections as Data: Part to Whole* initiative in 2018 also sparked this work (Padilla 2019, 2023). Our participation in the Fantastic Futures international conferences (AI4LAM), starting in 2019, provided a network of colleagues and partners interested in applying AI/machine learning processes to library, archive, and museum collections.

Ethical considerations

The University of California, Berkeley, library released their Responsible Access Workflows for digitization projects in 2020, which provides workflows covering four key law and policy areas relevant for digital collections: copyright, contracts, privacy, and ethics. Of particular interest for this project was the ethics workflow, which prompts staff to consider whether unfettered digital (or analog) access could result in the harm or exploitation of people, resources, or knowledge. If the answer is yes or uncertain, then the workflow calls for: reference to professional and community standards, community engagement, and adapting local policies to better support ethical engagement. This workflow helped guide our risk assessment for the records being digitized and aligned well with our plan to form a community advisory group to consider ethical access to these materials.

When writing the JACS grant proposal, we recognized that these records were potentially sensitive and would require additional evaluation prior to being released publicly. The context of the original data collection in WWII raised ethical concerns, as the information was taken under duress and without consent, and documented forcibly relocated and incarcerated individuals. Additionally, the original forms contain sensitive information, including personally identifiable information (PII) such as Social Security numbers, as well as religious affiliations, health conditions, work history, family relationships, hobbies and personal interests.

Implementing an automated transcription tool was useful within the context of this project because it helped us resolve some of the ethical concerns posed by putting these sensitive documents on platforms such as

Zooniverse for a public transcription project. In the interest of maintaining the relative privacy and security of the materials during the AI implementation stage, we developed a workflow with Doxie.AI to transfer digital images securely via Box, with extracted data deposited in a private Github repository for further evaluation and editing by library staff. Doxie.AI customized their machine learning pipeline in ways which further aligned with privacy and ethical concerns; for example, we decided to automatically redact Social Security and Alien Registration numbers from the dataset so we would in no way collect or aggregate this data at any point in the project. It is worth noting here that none of the data produced from this project was used by Doxie.AI towards their larger corpus of training data.

AI models are often the topic of controversy because of bias that is sometimes inherent in the data on which they are trained. This bias is often seen in areas such as facial recognition and text generation. For this project, we did not have any such use cases. The primary use case was OCR and the only language was English, so we were not affected by the above mentioned biases. One aspect that did present itself was the OCR of Japanese first and last names. OCR models often use general language understanding to boost their accuracy, and such data often consists of English names. Doxie.AI developed special post processing of names and places using custom dictionaries in order to accurately transcribe this information.

The library recognized that while these records were of tremendous research value, digitizing and providing access to them without adequate context and community input constituted poor stewardship. We worked closely with Densho, our community memory organization partner, and with our UCB library scholarly communications team to host a community advisory group meeting in August 2022. Our community advisors spanned a wide range of geographic and demographic backgrounds within the Japanese American community, which included former internees and descendants, activists, artists, community historians, writers, and mental health professionals in collaboration with librarians and curators. While we intended to hold this meeting in-person, after several Covid-19 surge-related cancellations, we decided to meet virtually to protect the health and safety of all participants. The project team's investment in creating a number of resources to inform participants of the project context and goals, as well as the library's ethics workflows and access policies, led to a very successful virtual community engagement. Abundant documentation was produced by Bancroft staff, UCB library colleagues in the scholarly communications office, and our community partner organization Densho to guide the day-long series of small- and large-group discussions. Throughout the meeting, participants were asked to weigh the benefits of public access to the data with the risks of potential harm to community members.

During the meeting, a general consensus emerged from the advisory group that the research and community value of the information in these materials outweighed the potential harms of opening the records for research, but that any release should be accompanied by adequate context and description to explain those circumstances under which it was gathered. Some logistical questions remain unsolved. For example, whether and how to introduce a more layered, multi-level approach to access was not decided,

and is dependent in part on staffing capacity and the available technical mechanisms supported by the IT infrastructure of the broader UCB library. Further follow up with library stakeholders and key decision-makers is needed before we can propose and implement an ethical access plan for both the digitized records and dataset. Until a final decision is made, both sets of resources will remain restricted from public access.

Who is affected by this project?

Many people and groups are and will be affected by this project. First, and most importantly, these archival records are not openly available to community members, so if and when we are able to make the digitized forms and dataset available, it will have a huge impact on the communities affected by the forced incarceration of Japanese Americans during WWII. Many have never seen the forms or been able to analyze the data that was omitted or generalized in the WWII-era punch cards. The details of work histories, education, hobbies, and other particulars of people's lives will finally be available to their families. Secondly, these records will provide important data to researchers looking at historical patterns represented in this massive demographic dataset. We hope this new information will provide new insights and, in combination with other datasets such as Densho's Name Registry and the Final Accountability Roster (FAR) records, provide a bigger picture of the impacts of this period in history (Densho "Name Registry" and "Final Accountability Roster" 2024). On the library level, we plan to provide access to this data as a collection, help users find and access it, and, as much as possible, support computational services around this data. This will largely impact our IT and research support services groups, who need to provide the infrastructure and tools to support these services, and as well as the experts in digital scholarship service areas who will support users.

Lessons learned and future work

Success in this project came from leveraging partnerships to combine technical expertise with content and domain expertise, as well as community knowledge. Responsible implementation of AI in this context relied upon these different knowledge communities to collaboratively develop a machine learning pipeline informed by considerations of privacy and ethics, and to apply an ethical framework for co-curation of the various digital resources produced by the project. Ethical implementation of AI should be iterative and collaborative, guided by clear policies and ethical frameworks, and informed by community engagement and input.

We learned that there is no one-size-fits-all approach to applying AI responsibly in an archival context. This is in large part related to the unique characteristics and contexts of discrete archival collections. Not all material within an archival repository lends itself readily to these tools in their current state of development, and selecting material that would benefit most from this process, especially given staffing or resource constraints, depends on a variety of factors. When adopting AI tools for mass transcription or data extraction, we have found it important to consider the structure and content of the material (i.e. does it already contain structured data), the size or quantity of materials in the collection, the consistency of the

physical records, and the risk of harm posed to record creators or subjects in making particular collections available online and accessible for computational research.

The Bancroft Library is still exploring the extent of our role in creating and providing access to “research-ready” data (i.e. data that is machine readable, queryable, maintains original structure, annotated, and linked to other data on the same topic data), and how AI/machine learning can be leveraged to help us do this work efficiently and economically. More resources are needed to simply digitize our collections and the additional cost of extracting research ready data strains already tight resources. Is creating clean and usable data that does not require significant additional work by the researcher enough? This project provided one tangible case study for how we might implement this in the future for similar materials but we need to improve cost models, potentially by partnering with researchers and community members to improve upon the machine-generated datasets. In this project, the AI/machine learning costs added over 50% on top of our usual digitization costs, and that is a big lift going forward as we still struggle to find funding for that work alone.

Our ethical work on this project is still not complete. We anticipate future work will include additional consultation with community members and other stakeholders in alignment with our ethical guidelines. Technology has given us an opportunity to provide much needed information to individuals and families affected by the Japanese American incarceration during World War II. We want to provide that information as thoughtfully and ethically as possible, guided by our community partners and individuals who were impacted by that traumatic experience and by the ethical practices that are developing and taking shape in our field.

Documentation

- 6th Computational Archival Science (CAS) Workshop: *Using AI/Machine Learning to Extract Data from Japanese American Confinement Records*
- University of California Berkeley Library: *Responsible Access Workflows*
- The Bancroft Library: *JACS6 Community Advisory Group Meeting Participant Packet*

Acknowledgements

Funding for this project was made possible, in part, by grants from the **U.S. Department of the Interior, National Park Service, Japanese American Confinement Sites Grant Program, and The Henri and Tomoye Takahashi Charitable Foundation.**

The research case study was developed as part of an **IMLS-funded Responsible AI** project, through grant number **LG-252307-OLS-22.**

Competing Interests

The authors declare that they have no competing interests.

References

- Adams, Nick. 2017. "From Static Archive to Research-Ready Database." *Library of Congress: IMPACT symposium*, October 30, 2017. <https://youtu.be/Ap5NHCujAVA>.
- Bancroft Library. 2019. "Japanese American Evacuation and Resettlement Records." The Online Archive of California. 2019. <https://oac.cdlib.org/findaid/ark:/13030/tf5j49n8kh>.
- Bancroft Library. 2022. "JACS6 Community Advisory Group Meeting Participant Packet." Accessed November 10, 2023. <https://docs.google.com/document/d/1idg1OyL12GouiSypbLz7UMsL4Bu4b21VuC3Jbq0e9b8/edit?usp=sharing>.
- Berkeley Center for Interdisciplinary Critical Inquiry. n.d. "Summer Minor in Digital Humanities." Accessed February 21, 2024. <https://cici.berkeley.edu/programs-and-initiatives/digital-humanities>.
- Collections as Data Facets. n.d. "#HackFSM." Accessed February 21, 2024. <https://collectionsasdata.github.io/facet9>.
- Densho. n.d. "Preserving Japanese American stories of the past for the generations of tomorrow." Accessed February 21, 2024. <https://densho.org>.
- . n.d. "Densho Name Registry." Accessed February 21, 2024. <https://ddr.densho.org/names>.
- . n.d. "Final Accountability Roster." Accessed February 21, 2024. https://encyclopedia.densho.org/Final_Accountability_Roster.
- Doolittle, Austin, Cameron Ford, Vijay Singh, and Tracey Tan. 2020. "BugTrAP: (Bug Transcription and Annotation Pipeline)." UC Berkeley School of Information, MIDS Capstone Project Fall 2020. <https://www.ischool.berkeley.edu/projects/2020/bugtrap>.
- Doxie.AI. n.d. "Doxie.AI." Accessed February 21, 2024. <http://doxie.ai/#>.
- Friedman, Marissa, Cameron Ford, Mary Elings, Vijay Singh, and Tracey Tan. 2021. "Using AI/Machine Learning to Extract Data from Japanese American Confinement Records." In *IEEE BigData'21 Computational Archival Science: digital records in the age of big data workshop proceedings*, Virtual, December 17, 2021. <https://doi.org/10.1109/BigData52589.2021.9672076>.
- National Archives. n.d. "[Japanese-American Internee Data File], 1942 – 1946." Access to Archival Databases (AAD). Accessed February 21, 2024. <https://aad.archives.gov/aad/fielded-search.jsp?dt=3099&tf=F&cat=all>.
- Padilla, Thomas, Laurie Allen, Hannah Frost, Sara Potvin, Elizabeth Russey Roke, and Stewart Varner. 2019. "Always Already Computational: Collections as Data: Final Report." DigitalCommons@University of Nebraska – Lincoln. <https://digitalcommons.unl.edu/scholcom/181>.
- Padilla, Thomas, Hannah Scates Kettler, and Yasmeeen Shorish. 2023. "Collections as Data: Part to Whole Final Report." Zenodo. <https://doi.org/10.5281/zenodo.10161976>.
- University of California Regents. 2021. "Berkeley Library, University of California, ResponsibleAccessWorkflows_PUBLIC_CC-BY-NC-4.0." Accessed November 10, 2023. <https://docs.google.com/presentation/d/1V66PGplq9xqXxdvngpD3rkAMolw2hlyVVDS4lv4VFOM/edit>.