



COLLEGE OF
INFORMATION
STUDIES

Organization: Digital Curation Innovation Center (DCIC), U. of Maryland iSchool
Project title: Automating the Detection of Personally Identifiable Information (PII)

Primary mentors: Richard Marciano, Professor
 Greg Jansen, Research Software Architect

Digital Curation Innovation Center (DCIC)
 University of Maryland's iSchool

Description	This projects looks at computational treatments of WWII archival collections and focuses on automating the detection of PII in historical index cards.
Problems	Computational challenges will be explored including: <ul style="list-style-type: none"> • Data creation, manipulation, analysis, and visualization • Designing and constructing computational models • Computer programming, developing modular computational solutions, and troubleshooting and debugging
Techniques	<ul style="list-style-type: none"> • Metadata extraction techniques: Named Entity Recognition (NER) • Developing, refining, and testing PII detection algorithms • Integration of name registries and linking.
Data	We will use 10,000 records of the WWII Japanese-American Incarceration Camp card series 52 (from NARA's Record Group 210: Records of the War Relocation Authority). This series has not yet been released to the public and includes index cards containing information about internal security cases for the ten incarceration centers in which Japanese-Americans were imprisoned during World War II. As of Nov. 1, 2018, the DCIC has received the green light to work with community groups to help disseminate this vital information.
Outcome	The project will partner with Densho.org in Seattle to work on processing and releasing these records to the public using community-vetted and approved access policies.