

2020

Hawai'i Annual Code Challenge (HACC)

Challenge Title	Gamification of Crowdsourcing Indexing
Department / Organization	Hawai'i State Archives
Subject Matter Contact	Adam Jansen, PhD
SME Phone & Email	808-586-0310
The Challenge	
Describe situation to be solved	<p>The State Archives has a large collection of historic records that can assist individuals in connecting with their personal history, heritage and culture. Unfortunately, many of these records are contained within larger collections of poorly identified record sets (i.e. Land Records, Vital Records) that lack individual record descriptions. Imagine trying to find the one record regarding your great-great grandparents amongst 100,000 other unrelated records. Most researchers do not have the time, patience or temperament to search through that many records on the <i>off chance</i> that they may find that one record they are looking for. To provide better end-user access to these records, individual indexing of the records is a necessity. To increase community interest in conducting remote transcription of these records, gamifying (applying game design elements) to the index process will provide incentive (through leaderboards) and fun (through virtual milestones and awards) to incentivize the volunteers to produce more indexed 'product'.</p>
Preconditions <i>(How does it work now)</i>	<ol style="list-style-type: none">1. Public researchers attempt to locate relevant records based on meta-descriptions of the bulk record set (i.e. Land Records, marriage records)2. Contents of that record set may, or may not, contain a relevant record3. Researcher looks through one folder at a time, then one page at a time to determine whether or not record is relevant to their line of inquiry.4. Researcher repeats process until relevant record is located, should one exist.
Assumptions/Issues <i>(list any conditions that could impact solution)</i>	<ol style="list-style-type: none">1. Interface needs to be accessible online2. Cannot require commercial/paid software (i.e. open source solutions preferred over commercial off the shelf products)3. As each record set is different, needs an interface to 'set up' an indexing project that specifies order, data types and names of data elements for each record set to be indexed (or similar ability to use and upload configured spreadsheets)

<p>Current Approach <i>(how is situation currently being handled)</i></p>	<ol style="list-style-type: none"> 1. Volunteers come to archives to access scanned versions of records 2. Using local laptops, volunteers access one scanned image at a time 3. Volunteers transcribe pre-determined data elements from the scanned image into spreadsheet, with each scan match a row in the spreadsheet 4. A second volunteer will validate the information entered by the first. 5. Upon completion of a set of records, the validated spreadsheet is sent to the technical team to load index and images together into online database for searching and discoverability.
<p>Users <i>(Who would use the application - employees or constituents or both? How many users would there be?)</i></p>	<ol style="list-style-type: none"> 1. Archive Employees 2. Government Legislature and Executive Researchers 3. Historians 4. Researchers 5. Real Estate professions or lawyers 6. Curious Public
<p>Business Rules</p>	<ol style="list-style-type: none"> 1. Method for importing required metadata fields to be indexed for each record set 2. Ability to match up an image with an index entry – NOTE: some records may contain multiple, separate index lines for each image. 3. Export indexed information in consistently formatted, text-based file format 4. Track which volunteer indexed which records 5. Allow admins to assign point value for each line of indexing based on number of fields and complexity of record 6. Display point totals to volunteers with dynamically updated leaderboard of ‘top score indexers’
<p>Special Requirements</p>	<p>Community participation to transcribe identified metadata elements from of sets of like records.</p> <p>Indexed record workflow allows for second volunteer validation of indexed information to assure accuracy of transcription</p>
<p>Technical Platforms <i>(in use or desired to be used)</i></p>	<ol style="list-style-type: none"> 1. Archives use Java, Enterprise DB/PostgreSQL 2. Strong preference to use of open source tools
<p>Data set to be used or collected</p>	<p>Photographs, Passenger Manifest Records, Land Records</p>
<p>Data set calculations or reporting needs</p>	<p>Not applicable</p>
<p>Solution Road Map</p>	

Basic Flow <i>(steps of user action/system response)</i>	<ol style="list-style-type: none"> 1. End users signs on to web app 2. List of available record sets needing transcription appears 3. User selects record set of interest to them 4. System presents scanned image of record along with dialog/text boxes soliciting specific data elements from record 5. For each box that the user populates with data, they are given points 6. System tracks point progress to provide 'badge', virtual awards, and point leader progress
Goal of Solution	<p>Department doesn't have the time and resources to index important metadata elements off of records. By allowing crowdsourced indexing of the most important genealogy and land rights records at the State Archives, discoverability of relevant records is greatly increased – allow the Archives to match up records to individual researchers. Scholars, genealogists, schools, community, and industry will have an opportunity to contribute their expertise in transcribing various historic record sources from throughout the history of Hawaii from Kingdom to current. Adding a gamification aspect to the indexing adds an element of accomplishment and friendly competition to incentivize volunteer indexers to transcribe records. Once transcribed, access to these important records can then be provided in a keyword searchable online repository to allow interested individuals to conduct genealogical and land rights research from the comforts of their home, anywhere, anytime, for any reason.</p>
Business Value <i>(potential financial or time savings)</i>	<ul style="list-style-type: none"> • Disseminate information from our collection of historic records and photographs • Compliance with public records laws • Reduction of onsite researchers by providing self-service access • Experts on a location/person/subject/event can contribute their knowledge to the indexed data in order to increase discoverability • Users able to access relevant information/records/persons filed in records sets that may not be an intuitive source of information, or that may be extremely voluminous amounts of information to dig through to find one hit (needle in a haystack syndrome).
Success Scenario <i>(how you know a solution is working)</i>	<p>Users able to use Hawaii State Digital Archives Catalog to perform keyword/name searches and have a results list returned that contain the relevant identified records</p>
To be completed by the HACC Planning Committee	
Community/Industry Data Available	<p>Not applicable</p>
Potential Community/Industry Co-Sponsors	<p>Not applicable</p>