Layering Historical Maps and Census Data for Interactive Visualizations in HistoryForge

Lovell, Claire; South Central Regional Library Council; Ithaca, New York, U.S.A.
Kibbee, Robert; The History Center in Tompkins County; Ithaca, New York, U.S.A.

Abstract

HistoryForge (https://historyforge.net) is a web application that combines information from U.S. Census records, historical maps, and other records in an interactive framework of human and spatial relationships that illustrate what communities looked like and how they evolved over time. It generates an environment that invites a study of local history at the levels of neighborhood, family, and individual. HistoryForge is being developed using open source software so that any community can adopt it to explore their own local history and add archival material. This paper will describe the project's development, growing potential for enriching records with archival material, and its current implementation in four different communities. The rapid development of the last year has been supported by a two-year grant from the Public Engagement with Historical Records from the National Historical Publications and Records Commission of the National Archives.

Motivation

The History Center in Tompkins County (THC), like most history centers and libraries with historical collections and archives, has been faced with the problem of how to provide compelling access to their archival material. As digitizing projects generate ever larger and more diverse digital collections, archives are challenged to display materials in online contexts that engage directly with their users. The ubiquitous gallery view, in which digital collections are simply presented as a list of objects, offers minimal interactivity or context. Often, this model for display overwhelms the user without inspiring deep exploration.

THC has developed a unique approach, motivated by the desire to display their archival collection in a compelling historical context and a commitment to design an interactive environment with open source software adoptable by any community. THC wanted to create this environment using two extremely rich but underutilized sources of primary historical material: the records from the manuscripts of the U.S. Census Bureau and the maps from the Sanborn Fire Insurance Atlases.

Problem

The larger problem is how to engage a community with its local history in a way that encourages historical narrative and an appreciation for the diverse communities that formed the modern city or village. THC framed this problem around discovery and access to primary documents, namely, the census records and historical map sets. As the project team developed an interactive online environment to do this, they realized the immense potential for developing record types and online record displays that could support additional archival content as well as user-donated content. This new environment would encourage an interactive engagement with archival material, as opposed to a passive viewing experience.

There were several barriers to implementing this approach. First was the daunting task of transcribing census records accurately and in a way that would translate the several dozen attributes into columns in a searchable database. The same process would also have to capture and normalize building addresses, so that each building and person could be located on both a modern "slippy map," such as Google Maps, and on the historical map overlays. Third was the difficulty of producing these geo-rectified map layers from digitized scans of individual maps and displaying them in the HistoryForge environment.

Another problem was how to design the environment in a way that it could be adopted by other communities. HistoryForge, although committed from the beginning to produce a project that any community could adopt, was inherently designed in an Ithaca-centric fashion. Thoughtful details built into the initial installation, such as the default of centering on Ithaca when geo-rectifying historic maps, caused inefficiencies and inconveniences for other communities.

Implementing HistoryForge in any new community would require some effort and flexibility, as each different city and village may have unique quirks in their physical development (such as street name changes and building demolitions), and the available primary sources may vary.

In particular, there was a significant development challenge to provide a discrete installation for each new community.

Approach

HistoryForge began with the simple idea of joining information from the manuscript censuses of the United States and the Sanborn Fire Insurance maps. The decennial censuses provide rich demographic information and the Sanborn maps provide extensive information about the built environment. The joined information would be presented in an interactive web environment.

After several proof of concept iterations, support from a local businessman funded a functional project designed by software developer David Furber from GORGES, LLC. A small but growing band of volunteers began transcribing and entering census data in April, 2014. This group has now entered over 80,000 individual census records, including 5 complete censuses for Ithaca, NY from 1900 to 1940.
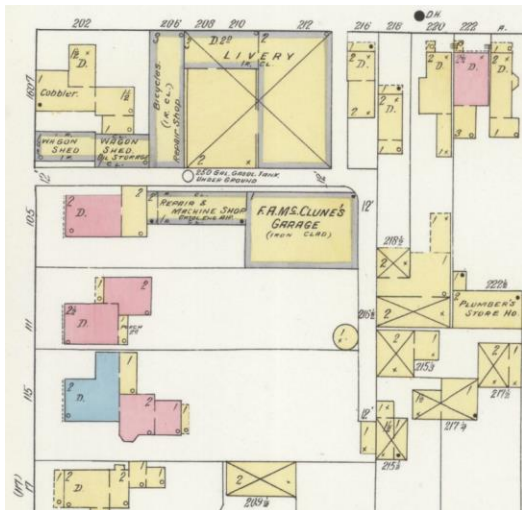
Figure 1. Part of a 1910 Sanborn Fire Insurance map of Ithaca digitized and made freely available by the Library of Congress, showing details such as street addresses, building material, building features, and business types.

Transcription sessions were popular social events that engaged the volunteers with one another and their shared local history. With the onset of COVID-19, the transcription, training and outreach all transitioned to a virtual platform, but the project's vision and output has remained the same.

A grant from the National Historical Public Records Commission (NHPRC) accelerated software development of an attractive, fully functional web environment. With that stage nearing completion, efforts shifted to implementation and maintenance models for testing partners so that new communities can adopt and run their own HistoryForge installations. This initiative has been amazingly successful and has resulted in three partners now actively engaged in creating their own HistoryForge environments: Auburn, NY, Elmira, NY, and Oberlin, OH.

The NHPRC grant supports a full-time Project Coordinator to provide training materials and extensive documentation for partners. The grant also provides the initial cost of installation and ongoing maintenance. Each testing partner has a grant-funded part-time local coordinator to manage volunteer recruitment, training and data entry. Additionally, this local coordinator supervises map-layer construction for their installation.

Technical Background

HistoryForge is a Ruby on Rails application running on a standard version of Unix Ubuntu. The application runs in the cloud on virtual servers supplied by Digital Ocean. Each community's installation runs on its own "Droplet," a virtual container. The project uses PostgreSQL, a standard open-source relational database, extended by PostGIS, also open-source, which provides support for additional spatial features.

The historical map layers are geo-rectified and mosaicked in MapWarper, another open-source tool designed and used to digitally rectify historical maps. Although there is a freely accessible online version, HistoryForge has installed MapWarper on its own server to maintain better control over map storage and export. Map files are stored in Amazon Web Services. Map layers are exported by MapWarper's map server via a Web Map Service (WMS) in order to run in the HistoryForge application.

Data Acquisition and Review

The HistoryForge platform includes data entry forms refined for the efficient and accurate input of highly-structured data from the decennial United States censuses of 1900 through 1940. While these records have already been transcribed for genealogical sites like FamilySearch and Ancestry.com, those transcriptions are incomplete and often inaccurate, and only available as individual records, not as records in a database table. Volunteers for HistoryForge use the freely accessible, digitized census images on FamilySearch, ultimately from the National Archives and Records Administration, to transcribe all the enumerated attributes directly into the HistoryForge forms.

The forms include available pop-up prompts for every field with instructions specific to that census year. These instructions have been designed and reworked by the Project Coordinator to be as concise and clear as possible. Census records tend to be quirky, so testing partners are encouraged to send in suggestions for these instructions.
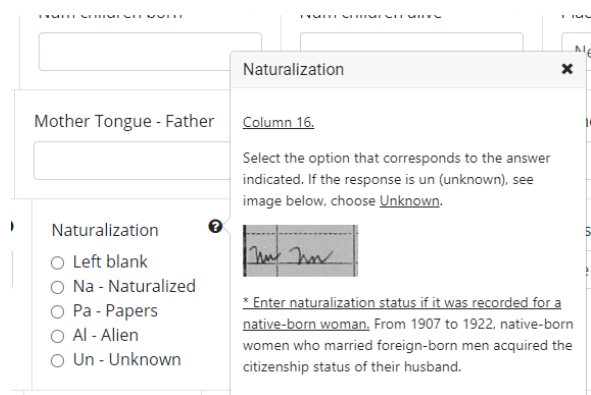

Figure 2. A partial screenshot of a pop-up window available for the 1910 census record form in HistoryForge, explaining the Naturalization field in detail and with a sample census entry shown.

The transcribed records are checked against other sources, such as city directories, and are reviewed by local coordinators before being added to the public database. Transcribers are instructed to write exactly what appears in the census record for the sake of data consistency. Each form in HistoryForge includes a free-text "Notes" field, where conflicting information gleaned from city directors or other sources should be written, which the local coordinators will review for possible corrections.

The forms include many time-saving defaults, such as "English" for the Language field, "New York" or "Ohio" as Place of Birth, etc. As they click "Save" on their finished record, transcribers are able to select from a short drop down menu to have the next record pre-filled for a person in the same family, in the same dwelling, on the same street, on the same page, or in the same enumeration district.

The resulting data set of census information is presented on each HistoryForge installation as a table with sorting and filtering capabilities, so the data can be examined and parsed interactively. Each piece of data available in the census year selected is available

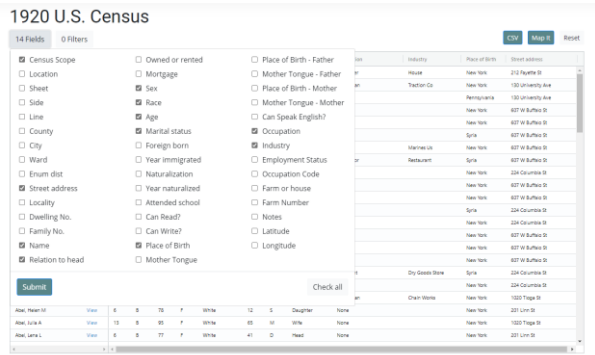to view, filter, export, and map. These fields change with each census year.

Figure 3. Screenshot of a table of transcribed census records with the Fields tab open. The Fields and the Filter buttons above the table have default selections that users can click on or off. The filter options are each tailored to the fields selected, so Side and Sex are both "one of either" options while Street address defaults to a "field contains text" option.

Census records are geo-addressed as they are entered. Transcribers enter the street address given on the census page and this information is turned into a structured address that can be sent to the Google GeoCoder, which provides a latitude and longitude for each building location. Local coordinators review these building locations and are able to easily adjust for precision using the modern street map or any of the historical map overlays uploaded to the installation, with the additional feature of an opacity slider. This is especially useful in cities that have experienced street number updates over the 20th century, causing the street address referenced in a census record to be somewhat or significantly different from the street address Google recognizes today. Certain historic maps, including Sanborn maps, show street numbers, so the building reviewer can cross-reference an older map to make adjustments.
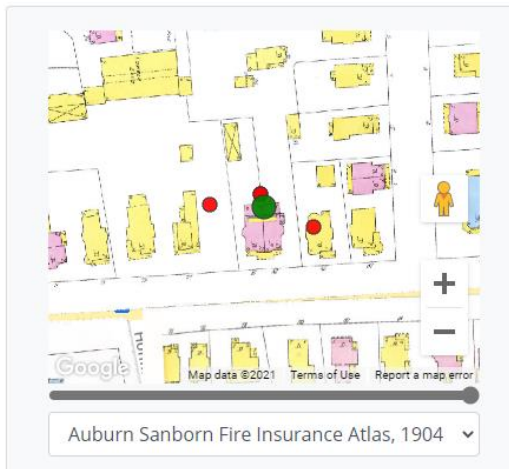

Figure 4. Partial screenshot of a building record page. The green dot represents the building being edited, while the red dots are recognized by HistoryForge as "Nearby Buildings," hyperlinked for ease of access. Reviewers can simply click and drag the green dot into place.

Each newly transcribed census record creates a person record and inherits a relationship with a building. Once the transcriber enters the street name into the form for a new census record, the form offers a drop-down list of existing buildings on that street. The transcriber can skip the drop-down list and instead click "Add building with address" for new buildings not yet inventoried. Local coordinators are responsible for reviewing any "unhoused" person records and "uninhabited" building records, as well as any records specifically marked by the initial transcriber as needing "investigation."

Maps and Mapping Data

Since every person record is tied to a building record and every building record includes latitude and longitude, HistoryForge is able to map all buildings and all listed residents in the map window. In the map window, users can experiment with layers of historic maps. The page of the web environment which displays all the buildings is named "The Forge" and is not tied to a particular census.

For the user, the exploratory process is to filter the records for a census year by one or more attributes of interest. If the user is interested in gaining an understanding of the Hungarian immigrant community in Ithaca in 1910, she can add a filter for Place of Birth or Father's Place of Birth as "Hungary." All the people meeting those criteria will be listed in a table that can be sorted or filtered again. Similar filters can be applied for race, occupation, unemployment status, home ownership, or any of the questions asked in a given census year. The data can be downloaded as a CSV that provides a basis for further analysis and, of course, unlimited application in other programs. The downloaded data, including the inherited longitude and latitude, can be easily used in a geographic information system (GIS). Most powerfully for average users, the filtered set of person records can be displayed as markers on HistoryForge's modern map by clicking on "Map It."

At that point, overlays of the historical maps are available and can be added to the map window ("The Forge"). The value of the historical layers is obvious in understanding places that have undergone substantial changes from Urban Renewal, changes in the industrial base, or other disruptions. Every historical layer has an opacity slider so that the historical topography and the modern one can be easily compared.

The historical map layers are geo-rectified ("warped") and mosaicked in MapWarper, an open-source tool installed on HistoryForge servers. Geo-rectification is the process of warping the map image onto a modern map, making it pannable, zoomable and most importantly, addressable. The separate plates of an atlas are individually geo-rectified and then merged into a seamless mosaic or overlay. The street addresses carefully captured from the census are geo-addressed and displayed on the modern Google map.

Map layers are exported from MapWarper via a WMS to run in the HistoryForge application, which uses React JS to display the maps and geo-addressed (through the Google Map GeoCoder) data points.

Initially project leaders had expected to digitize historic maps themselves and had anticipated this as a time-intensive and necessary step for other HistoryForge installations as well. However, the Library of Congress Geography and Map Division is actively digitizing their extensive collection of Sanborn Fire Insurance Maps and has made these digital files freely

available. All three testing partners are working on the 1910 census, so the maps of primary interest are from that time period and are therefore in the public domain.

The digitized maps expertly scanned by the Library of Congress are in excellent condition with minimal bleed-through and virtually no distortion, but the images do require modest editing before being used in HistoryForge. The MapWarper tool includes a cropping feature that can handle complex polygonal shapes, but compass roses, scales, publisher information, map titles, and plate numbers, which are often printed inside the map boundaries, must be removed in a photo editing software before uploading to MapWarper for geo-rectification. Map images are sometimes brightened as well, as this makes them easier to see in the HistoryForge map window.
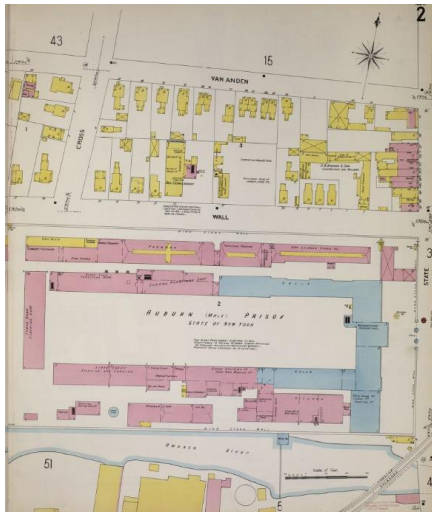

Figure 5. The scanned image of Plate 2 of the 1904 Sanborn Fire Insurance Map of Auburn, NY, which includes plate numbers and a compass rose.

Figure 6. Three separate plates, including Plate 2 seen in Figure 5, from the 1904 Sanborn Fire Insurance Map of Auburn were merged together in Adobe Photoshop, brightened, and edited. Editing and merging the maps in Photoshop offered more precision for irregularly shaped plates and streets.

While the Library of Congress has made available extremely high-resolution uncompressed TIFF files for the digitized maps, MapWarper's functionality is limited to JPG files of 20 MB or less. HistoryForge installations must strike a balance between map detail and file size, especially with regard to load time and the storage costs of large map mosaics.

Building a scaffold for archival materials

HistoryForge's structure invites an exploration of a community at a bird's eye view, but users immediately want to dig deeper into individual buildings and people to see a fuller picture of that time and space. To that end, the Program Manager and Project Coordinator, already working within a historical repository (The History Center), reached out to a multi-type library system with a well-established digital repository in order to build a scaffold for archival materials within HistoryForge. The library system, the South Central Regional Library Council, is one of nine library councils in New York State and a regional administrator for New York Heritage Digital Collections (https://www.nyheritage.org).

The building records in HistoryForge are authority records and make up the carefully managed inventory of buildings as reviewed by local coordinators. After consultation with the library system about NYHeritage's descriptive metadata standards, new forms were added to HistoryForge to accommodate photographs with structured metadata. These new records for archival materials are attached to the building authority records and accessible when a user clicks on a building address in the map window. The same intake forms will soon accommodate user-donated materials.

Authority records for residents are more challenging and still under development. An authority record for each person will allow for tracking individuals across time, linking to other data sources such as city directories, and adding archival material such as photos, obituaries, letters and clippings. Creating unique person records is not just a technical challenge, but will require significant effort to establish authority criteria and accurately connect records.

Creating a community of testing partners

From the outset, the project leaders envisioned HistoryForge as a web environment for exploring local history that any community could use. The core data resources are available to every community: the U.S. Census provides extensive information on people from every locality and Sanborn maps, which were created every eight to ten years from around 1870 to 1950 for most villages and cities of the United States, are widely available.

It was also clear from the outset that a recruitment model relying on "if you build it, they will come" would not be successful. The project was simple in concept but complex in implementation. As initially presented, a standalone installation of HistoryForge would require IT expertise beyond the reach of most historical societies and libraries. Successful installation would also require a dedicated project coordinator at each site to organize the volunteers, review records, and generally oversee the HistoryForge effort.

To address these problems the project team devised a model that would attract libraries and historical societies by enlisting them as "testing partners" within the NHPRC grant request. Special funding within the NHPRC grant was set aside for local

coordinators, thereby incentivizing organizations to become early adopters of this platform.

The three testing partners that joined between 2020 and 2021 benefitted from the NHPRC funding, which covered their local coordinator's staff time, as well as installation, hosting, and maintenance of HistoryForge. These early adopters also benefited from extra support and attention from the project leaders, with whom they have routine meetings and an open line of communication.

This is a symbiotic relationship, because as HistoryForge is implemented in more communities, this extended, personalized level of support may not be possible. The Project Coordinator uses the feedback provided by testing partners in their meetings and quarterly reports to better develop a comprehensive set of training guides and documentation so that new installations do not require such intensive oversight and assistance, thereby making the overall HistoryForge project more sustainable in the long term.

The testing partner model is further mutually beneficial because these early adopters bring new ideas and initiatives. The Auburn, NY, installation is especially interested in fleshing out the people, places, and relationships of a particular neighborhood. Within six months of their initial installation, they were awarded a local grant to hire interns to research and digitize archival materials in support of their HistoryForge environment. Their successes and challenges in the coming months will inform all other HistoryForge installations.

Results

Two years of intense development under the NHPRC grant have resulted in HistoryForge (https://historyforge.net), a successful web environment for exploring local history. The site provides new and compelling access to important but underutilized primary historical resources. The Ithaca HistoryForge and three testing partners have transcribed almost 100,000 individual census records, entering them into searchable databases, adding enormous value to data previously analyzable only in aggregate and for higher-level geographies (for these censuses, 1900-40, mostly at the county level or above). HistoryForge allows users to study local history at the granularity of individuals, families, and research-defined communities based on ethnicity, occupations, home ownership and other attributes.

Users can explore the spatial relationships of those individuals and communities with spatially-aware building records (and, by inheritance, person records). The relationships can be displayed on the modern topography or layers of historical maps. History is no longer limited to official geometries, such as wards, municipalities, towns and counties, but is now possible on the level of blocks and neighborhoods.

Archival materials can now enrich our understanding of the built environment by linking to the authority record for each building. The structure for similarly detail-rich person records is in development. The historical repositories participating in HistoryForge are taking advantage of the interactive web environment for access to archival material with additional digitization efforts.

The platform has been generalized for partner installation. HistoryForge now has three successful testing partnerships that will provide a model for additional partners and improve the documentation for future and even independent installations.

Since HistoryForge is open-source and available on GitHub, anyone can install it on their own servers. However, most historical repositories and cultural organizations will be attracted to the hosted site installation model, in which the Ithaca installation is replicated on another Digital Ocean droplet. HistoryForge also provides extensive documentation and training materials, as well as personalized training on digital platforms such as Zoom.

The testing partner experience: Auburn, NY

In Auburn, NY, the HistoryForge project was embraced by the public library, whereas the Elmira, NY, and Oberlin, OH, installations were both taken up by historical societies. Seymour Public Library District in Auburn has a well-established local history center with experience in other history-related digital projects, including newspaper digitization and born-digital oral history projects.

The library has been able to leverage its wide audience in recruiting volunteers, and the librarian serving as HistoryForge's local coordinator in Auburn already had supervisory and technology training experience, which have proven to be important staff skills for making HistoryForge successful.

The Auburn HistoryForge project has been significantly bolstered by a partnership with the Cayuga County Historian's Office, also located in Auburn. The County Historian's Office has provided invaluable knowledge, especially for identifying street name changes and geo-rectifying historic maps in a city whose original downtown was half demolished and rebuilt during Urban Renewal.

The HistoryForge project in Auburn has helped the two public organizations forge a strong collaborative relationship. Multiple staff members from the library and from the historian's office are members of the Auburn HistoryForge Advisory Committee, and have participated in the planning process, and have attended the monthly check-in meetings. These professional relationships will continue to be valuable in future community-wide endeavors.

Just a few months after installation, Seymour Library applied for and was awarded a digitization grant of $6,000. This grant will fund two interns who will research and select materials from both the library's historical archives and the county historian's collections to add to HistoryForge.

While all three testing partners started by transcribing the 1910 census, Auburn is also endeavoring to create a thorough, multi-faceted exploration of one particular neighborhood around Parker Street. The interns will perform research on the neighborhood's buildings, residents, and businesses to add rich descriptive detail and digitized images to the basic HistoryForge records.

The librarians in Auburn selected the Parker Street neighborhood for its diversity; while Auburn is home to many extant mansions and famous residents, the librarians are purposefully using HistoryForge to highlight underrepresented people and stories. The blend of census records and maps in HistoryForge perfectly serves this cause, as users can click "Map it" and see the clusters of "Mulatto" for race, "Ireland" for place of birth, or "Laborer" for occupation, among other details.

Figure 7. *Two views of the Parker Street neighborhood in 1910 on Auburn HistoryForge. On the left is the complete display of residences, each represented with a red dot. On the right is a filtered view, showing only buildings with non-white residents.*

Of note, volunteers in Ithaca began with in-person group transcription events long before the pandemic, and these group sessions survived the transition to virtual events. In contrast, the Auburn volunteers, who started their work during the pandemic, have shown little interest in group transcription via Zoom and primarily work independently. HistoryForge is developing as a flexible project, so that testing partners can pursue the approach that interests them and work with their volunteer base as they prefer. Documentation provided to new partners now reflects best practices learned from all of the different approaches.

Conclusion

HistoryForge and similar interactive programs offer elements of public participation, utilization of public records re-organized as searchable databases, and spatial visualizations based on historic mapping. These provide a new and powerful framework for displaying many archival materials, and a far more compelling presentation than the typical gallery views, inventories, or simple spatial indices. Materials in this framework are presented as elements of a dynamic, continuously developing historical narrative.

Author Biography

Claire Lovell is the Digital Services Librarian at the South Central Regional Library Council, a multi-type library system in Ithaca, NY. She is also an active volunteer in the Auburn, NY, HistoryForge project.

Bob Kibbee was the Map and Geospatial Information Librarian at Cornell until 2010. Most of his career at Cornell was as a reference librarian specializing in census data. Bob has a strong interest in cartography and historical geography.