7-25-2022

Creating a Public Space for Georeferencing Sanborn Maps: A Louisiana Case Study

Adam Cox

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Part of the Archival Science Commons, and the Geographic Information Sciences Commons

Recommended Citation

Cox, Adam, "Creating a Public Space for Georeferencing Sanborn Maps: A Louisiana Case Study" (2022). LSU Master's Theses. 5641.
https://digitalcommons.lsu.edu/gradschool_theses/5641

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
CREATING A PUBLIC SPACE FOR GEOREFERENCING
SANBORN MAPS: A LOUISIANA CASE STUDY

A Thesis
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Geography and Anthropology

by
Adam Christopher Franklin Cox
B.A., Lawrence University, 2011
August 2022
To Grampa, who would have really liked this project.
Acknowledgments

I would like to thank my advisor Andrew Sluyter and the rest of my thesis committee for their guidance throughout this project. Also, thank you to the G&A and SLIS office staffs, and graduate school administrators for helping me through the process. This project owes much to Andras Nagy for his dogged testing and feedback via GitHub. For collaboration and support in one way or another, many thanks to Bert Spaan (Allmaps), John McEwen (SCAUG), Eliot Jordan (GeoBlacklight, IIIF Maps), Julie Stoner (LOC), Sydney Shivers and Sam Buckley (APA), Elisabeth Stancioff (LaURISA), Becky Smith (HNOC), Florian Hoedt, Toni Schoenbuchner, and Dave Kennewell (GeoNode), and Jeff Meyer (OpenHistoricalMap). Too many other people to name here sent me messages of interest and encouragement at some point during this work, each one of which, however small, meant a lot to me. Finally, for patiently listening to me talk about this project for hours and for supporting me throughout, thank you, Nicole.

Most of all, a huge debt of gratitude is owed to all of the crowdsourcing participants who got interested in this project (especially WallyKitty)… from me, of course, but also from future explorers who will get to discover all of the Louisiana history in these Sanborn map layers. Thank you all!
# Table of Contents

Acknowledgments .................................................................................................................. iii
List of Tables ........................................................................................................................ vi
List of Figures ........................................................................................................................ vii
List of Abbreviations ............................................................................................................ viii
Abstract ................................................................................................................................... ix

Chapter 1. Introduction ........................................................................................................ 1

Chapter 2. Research Design ............................................................................................... 5
  2.1. Crowdsourced Georeferencing in Context ................................................................. 5
  2.2. Software Approach ...................................................................................................... 15
  2.3. Designing a Pilot Project ............................................................................................ 24

Chapter 3. Using Sanborn Maps ........................................................................................ 31
  3.1. History of the Sanborn Map Company ...................................................................... 31
  3.2. Accessing the Library of Congress Collection .......................................................... 34
  3.3. Georeferencing Sanborn Maps ................................................................................. 39

Chapter 4. Building the Platform ....................................................................................... 46
  4.1. Overview of the Development Process ................................................................... 46
  4.2. The Georeference Extension ..................................................................................... 48
  4.3. The LOC Insurance Maps App ................................................................................. 63
  4.4. Documentation .......................................................................................................... 70
  4.5. Notes on the Development Process ......................................................................... 70

Chapter 5. Pilot Project Results .......................................................................................... 72
  5.1. Overview .................................................................................................................. 72
  5.2. Outreach Strategies .................................................................................................. 73
  5.3. Georeferencing Activity ............................................................................................. 78
  5.4. User Survey Questionnaire ...................................................................................... 83

Chapter 6. Discussion .......................................................................................................... 86
  6.1. Considering the Platform ......................................................................................... 86
  6.2. Crowd Activity .......................................................................................................... 97

Chapter 7. Conclusion ........................................................................................................ 103

Appendix A. IRB Materials ................................................................................................. 108
  Determination Letter ....................................................................................................... 108
  Informed Consent Form ................................................................................................. 109

Appendix B. User Survey Questionnaire .......................................................................... 111
  User Questionnaire for LaHMG .................................................................................... 111

Appendix C. Session Data Examples ............................................................................... 114
  PrepSession.data .......................................................................................................... 114
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeorefSession.data</td>
<td>114</td>
</tr>
<tr>
<td>TrimSession.data</td>
<td>115</td>
</tr>
<tr>
<td>Note</td>
<td>115</td>
</tr>
<tr>
<td>References</td>
<td>117</td>
</tr>
<tr>
<td>Vita</td>
<td>123</td>
</tr>
</tbody>
</table>
List of Tables

Table 4.1. Georeferencing Status thesaurus keywords.................................................................53
Table 5.1. Chronological list of prominent outreach efforts........................................................77
Table 5.2. Users that performed one or more task, ordered by date joined................................83
Table 5.3. Summary of user-reported interests...............................................................................85
List of Figures

Figure 2.1. A user participation matrix for participatory archives.........................................................9
Figure 3.1. Origins of today’s digital Sanborn map collections.................................................................37
Figure 3.2. Breakdown of a Sanborn map sheet—Alexandria, La., 1900, sheet 1.................................41
Figure 4.1. Layered architecture of the georeference extension and LaHMG app.................................46
Figure 4.2. High-level diagram of the platform structure........................................................................47
Figure 4.3. The preparation process........................................................................................................55
Figure 4.4. The georeferencing interface.................................................................................................57
Figure 4.5. Architecture of the live georeferencing preview.................................................................58
Figure 4.6. The trim interface................................................................................................................59
Figure 4.7. The Georeference tab............................................................................................................61
Figure 4.8. Data model for the georeference extension...........................................................................62
Figure 4.9. The volume summary page for Lake Charles, 1903.............................................................66
Figure 4.10. Data model for the LOC Insurance Maps app....................................................................69
Figure 5.1. Half-page flyer used for outreach..........................................................................................73
Figure 5.2. Timeline showing new and total registrations......................................................................76
Figure 5.3. Full timeline of all georeferencing activity............................................................................78
Figure 5.4. Duration of preparation sessions...........................................................................................79
Figure 5.5. Duration of georeferencing sessions....................................................................................80
Figure 5.6. Duration of trim sessions......................................................................................................81
Figure 5.7. Overall session timeline for New Orleans, 1885, Vol. 2....................................................82
Figure 6.1. Showing all sessions sorted by user—WallyKitty or anyone else........................................100
List of Abbreviations

API – Application Programming Interface, a machine-readable method for interacting with software or databases.

CMS – Content Management System.

COG – Cloud-Optimized GeoTIFF, a geospatial data format used for storing and serving raster images.

GCP – Ground Control Point, a linkage between a point in an image and a location in a GIS coordinate reference system.

GDAL – Geospatial Data Abstraction Library, a powerful and ubiquitous low-level library for handling raster data.

GeoJSON – A specialized JSON structure for storing geographic data.

GeoTIFF – An augmented image file embedded with spatial coordinates.

GIS – Geographic Information System(s).

JSON – JavaScript Object Notation, a flexible structure for storing text or numerical data.

LOC – Library of Congress.

OGC – Open Geospatial Consortium, an international organization that promotes and creates open geospatial software and data standards.

VRT – Virtual Raster, a file format native to GDAL that stores pointers to one or more raster datasets.

WMS – Web Map Service, an Open Geospatial Consortium standard protocol for serving and consuming geographically referenced images, like georeferenced historical maps.
Abstract

As institutional archives digitize their historical map collections and make them publicly available online, new methods for engaging with these materials emerge. Georeferencing the maps transforms their content from static images to dynamic map overlays, and allows for the extraction of geographic data like building footprints or place name coordinates. Many organizations have turned to crowdsourcing to georeference their large holdings, and this thesis approaches crowdsourced georeferencing from the perspective of participatory heritage, taking much inspiration from the idea of the archival commons. To test these ideas, a new extension was created for GeoNode—an open source geospatial content management system—that allows users to georeference map documents in a web browser. Further augmentation facilitated direct ingestion of digital content from the Library of Congress Sanborn Map collection, and a pilot project was conducted to engage the public in georeferencing maps of towns and cities across Louisiana. By the end of the project, 66 participants—from within Louisiana and without—had georeferenced all or a portion of 267 different Sanborn map volumes, creating over 1,500 new layers. These layers combine to create mosaics of 138 different communities across the state, including comprehensive coverage of the city of New Orleans in the years between 1885 and 1893—an especially valuable dataset in its own right. Seamless mosaics will be made from these layers and published via the LSU Atlas data portal for long-term public access. This experience led to new ideas for how to better engage citizens with historical maps of their communities, while the underlying construction of the georeferencing system itself provided insight into how users participated in the work. Ultimately, this thesis lays a conceptual foundation for future efforts of a similar nature, whether they pursue exactly the same technological approach or not.
Chapter 1. Introduction

Over the past two decades, GLAM institutions (galleries, libraries, archives, and museums) have embraced the opportunity to curate and disseminate their historical map collections online, resulting in publicly-accessible databases of scanned maps that number in the hundreds of thousands. This is especially relevant for historical geography, where the “spatial turn” has created new demand for geospatial analysis of historical materials, and maps are the obvious bedrock of such research (Knowles 2008; Bidney and Piekielek 2018). However, while putting map collections online does greatly expand access to them, the maps can be transformed further into more useful and accessible content for research. This thesis concerns one such transformation process called “georeferencing,” wherein the features on a scanned map are linked to real-world latitude/longitude coordinates, and the digital file is processed to embed these coordinates. The transformed result is, typically, a geospatial “raster” layer that researchers can combine with other spatial data for use in desktop software like QGIS or ArcGIS, or incorporate into interactive web maps through open data exchange standards.

Naturally, archival institutions have long understood the advantages that a fully georeferenced map collection can provide, but transforming an entire collection is typically far too much manual labor to be carried out solely by the staff of a single organization. To solve this issue some institutions have turned to crowdsourcing campaigns; by allowing members of the public to engage in the work thousands of maps can be processed in a short amount of time. To carry out public projects like this, institutions typically implement web platforms that have been built specifically to facilitate georeferencing.
There are not many such platforms available for use, however, so the first portion of the thesis will consider the existing options and then design and create a new one built from the content management system GeoNode,¹ taking inspiration from participatory heritage theory such as Alexandra Eveleigh’s user participation matrix for participatory archives (Eveleigh 2014), and the *archival commons* (Anderson and Allen 2009). The second portion of this research carries out a public pilot project with the new platform by tailoring it to the online Sanborn Maps collection at the Library of Congress (LOC).²

This research will be meaningful for a number of reasons, not least of which is simply that a new georeferencing web platform should be a welcome addition to a software realm with relatively few offerings. Beyond that, the approach should help to decentralize access to crowdsourced georeferencing initiatives because the ability to independently build off of the existing LOC Sanborn map collection will allow smaller entities, like rural historical societies, to use technological approaches usually only available to large organizations with plentiful resources. Finally, while participant feedback from the pilot project will provide direction for next steps in the development of the platform, the georeferenced Sanborn maps results will stand alone as a cohesive set of historical maps layers that will be publicly disseminated online through the LSU Atlas geospatial data portal for long-term access.³

To begin presenting the work here, Chapter 2 will lay out the research design for the project, first elaborating on the concept of crowdsourced georeferencing and how it fits into the context of participatory archives, and then describing the most commonly used existing platforms. With this background, the preliminary ideas behind the software design will be

1 GeoNode, [https://geonode.org](https://geonode.org), last accessed July 7, 2022.
presented. The chapter will conclude with an introduction to the scope and logistics of the pilot project, and elaborate on the reason Sanborn maps are a suitable choice for this application.

Chapter 3 goes much deeper into Sanborn maps, beginning with a history of the company and background on fire insurance mapping in the United States. Following that, the chapter provides a history of the Sanborn map archive, with special attention paid to how the LOC collection came to exist. An examination of the structure of Sanborn maps follows, and the chapter ends with a very simple overview of a potential workflow for georeferencing them—not just a single map, but how to approach the process as it applies to an entire volume.

Chapter 4 covers the actual design and creation of the georeferencing platform itself—the rationale behind its construction, the affordances it offers for participation, and the methods used to create it. The first part of the chapter focuses on the georeferencing capabilities of the system, while the second part describes the extra components were built specifically to reflect and build from the internal structure of Sanborn maps.

Chapter 5 presents the results of the public pilot project, with figures charting how many people used the system and how many maps were georeferenced (a lot!). This chapter also provides a breakdown of the various outreach efforts that were used to engage different members of the public, and a summary of the results from a simple user survey that was distributed at the end of the public participation period.

Chapter 6 discusses the results from the pilot project in two parts, the first focusing on the platform design itself—how well it embraced the idea of the archival commons, reflections on the soundness of the technological approach, and ideas for future improvements. The second section examines the nature of public participation during the pilot project, pointing out how
different categories of user activity became apparent, and identifying how to benefit from these types of participation in the future.

Chapter 7 concludes the document by recapping the overall objectives of this research and drawing findings from the results. While participants did georeference a large number of Sanborn maps by the end, it is most important to analyze which aspects of the platform best facilitated their work, and which ones hindered it. Examining these parts of the work reveals many ways to improve similar efforts in the future.

This thesis allowed me to bring together many interests that I have developed over the past few years. Open source geospatial software, georeferencing, historical maps, web design and programming, software architecture, public engagement, and even Sanborn maps—all of these elements have been lurking around my professional career, so it has been a pleasure to finally have the opportunity to combine them. I hope this enthusiasm shows through in the work, and I believe this thesis not only contains some original insights but has also produced geospatial resources that will be useful for years to come.
Chapter 2. Research Design

To set the stage for the work to be done, this chapter will begin with some background on the concept of georeferencing historical maps for research purposes, and then move into a discussion of how crowdsourced georeferencing can fit within other participatory heritage efforts. An examination of existing crowdsourced georeferencing platforms follows, specifically with regard to how they can inform new applications. Finally, the initial outline of the pilot project takes shape, including the introduction of the Sanborn maps collection and why it is a good fit for this research.

2.1. Crowdsourced Georeferencing in Context

2.1.1. Why Georeference Historical Maps?

Historical geography, as a field of study, began experimenting with geographic information systems (GIS) in the late 1980s (Miller 1986) and developed these research techniques through the 1990s (Mires 1993; Colten et al. 2003). In the early 2000s, Kelly Anne Knowles prefaced her edited volume Past Time, Past Place with a piece called “Introducing Historical GIS,” acknowledging a new focus on using digital spatial analysis tools for research in the field (Knowles 2002). This same sentiment was echoed in 2007 by Ian Gregory and Richard Healey (Gregory and Healey 2007). Others have more broadly described a “so-called spatial turn in the humanities,” and pointed out that the ever-increasing awareness and access to GIS technology has enabled this growing side of the discipline (Offen 2013; Harris, Rouse, and Bergeron 2010, 124).
However, as GIS becomes more deeply ingrained in historical geography, an inherent tension between the two emerges. This tension has its roots in the fact that traditionally, GIS research relies on a very quantitative approach, which does not always lend itself well to humanities research with qualitative complexities and nuance (Knowles 2008). GIS data uses exact coordinates, makes calculations with Euclidean geometry, and can run filters on specific data labels and tags. Thus, there is typically no accommodation for uncertainty or “fuzziness,” a common characteristic of humanities-related data (Gregory and Ell 2007, 39–40).

In acknowledgment of this, Harris, Rouse, and Bergeron (2010) proposed what they call a move toward a “Pareto GIS.” The Pareto principle, initially developed in the context of economics, has been extended to many other applications. Most generally, it posits that “for most events, 80% of the effects come from 20% of the causes” (Harris, Rouse, and Bergeron 2010, 129). To apply this principle to the design of a historical GIS is to recognize that perhaps 80 percent of the system’s value will come not from the less-used and more technical, analytical capabilities that GIS can provide, but from the 20 percent of utilities that are needed for basic display and aggregation of spatial data and historical maps. With this in mind, improving digital access, distribution, and integration of actual historical maps should be beneficial far beyond the more esoteric potential for advanced, geospatial analysis.

Naturally, as the web evolves so must web-based historical GIS. We are now well into the internet age of the “Semantic Web” (Berners-Lee et al. 2001), characterized by an emphasis on the creation of web platforms and content that can interact with each other and a de-emphasis on isolated applications. In the geospatial world, this has been nothing short of revolutionary, leading to the development of specifications by international bodies like the Open Geospatial
Consortium (OGC) that standardize geospatial information exchange between web-based providers and databases. Now, we have a “Geospatial Semantic Web” (Harris, Rouse, and Bergeron 2010), where government agencies, institutions, and private individuals alike can publish and consume geospatial data in any interactive web maps or desktop GIS software. Recently, the heritage world has also begun to engage with these sorts of web technologies, and Henriette Roued-Cunliffe’s book *Open Heritage Data* (2020) serves as a veritable call-to-action for heritage organizations to facilitate public engagement not by simply creating digital archives, but by exposing online collections through application programming interfaces (APIs) to allow machine-readable interaction, and programmatic access for amateur software developers. She considers the facilitation of this type of digital access imperative, claiming that “the consequences of not doing this would be a loss of relevance to society and a potential loss of revenue from both public and private funding sources” (Roued-Cunliffe 2020, 1).

Considering these themes with regard to historical maps, we can look at the evolution of digitization efforts undertaken by institutional archives. Initially, georeferencing scanned maps had nothing to do with digitization as maps received the same treatment as photographs or paintings; archivists scanned, cataloged, and uploaded the images to create a digital mirror of the physical archive. A 2015 survey of large institutional collections recorded impressive progress in these endeavors. For example, the British Library, New York Public Library, and David Rumsey Map Collection (a famously large private map collection) had scanned and uploaded 58,093, 38,908, and 38,336 of the maps in their collections, respectively (López 2016). Since then, of course, these numbers have increased. However, just scanning a map and putting it online does not embed spatial coordinates in the file, so it cannot be compared—either visually or
analytically—with other maps or spatial data. For that to be possible, these collections should be transformed by georeferencing each map.

In many cases, georeferencing has been automated (Herold et al. 2011; Fishburn, Davis, and Allord 2017; Piekielek 2017; Southall et al. 2017) or semi-automated (Burt et al. 2019; “NLS Map Georeferencer (Note)” 2017). Such automation typically relies on image recognition software that can detect specific locations within the scanned map, such as the four corners of the map’s border, and transforms the image by programmatically applying coordinates to these corners. This works well for historical map series with uniform styles, formats, and/or extents but for others with less regularity it does not (Mendt 2014; Burt et al. 2019). In the absence of automation, georeferencing must instead be performed manually by an individual, who will find points on the map—the corner of a building or intersection of two roads, for example—and then assign the latitude and longitude coordinates representing that feature in the real world. Desktop GIS applications like ArcGIS Pro or QGIS facilitate this process, making it a relatively straightforward point-and-click procedure. But how can an institution with tens (or hundreds!) of thousands of scanned maps georeference them all, if that process cannot be automated? One answer to this question is “crowdsourcing,” a citizen-participation data collection, creation, or curation strategy that, at this point, is well-used in the humanities (Hedges and Dunn 2018; Benoit III and Eveleigh 2019b).

2.1.2. A Framework for Crowdsourcing in Participatory Archives

To better understand crowdsourced georeferencing, we will begin by looking at it in the context of other heritage crowdsourcing activities. In “Crowding out the Archivist? Locating Crowdsourcing within the Broader Landscape of Participatory Archives” researcher Alexandra
Eveleigh laid out a User Participation Matrix (Figure 2.1) to represent the modes of user engagement across the variety of crowdsourcing efforts (Eveleigh 2014). Although she uses this framework as a way to understand the role of the professional archivist within each of these participation modes, the role of the archivist is not our main concern here. Instead, we will look at the characteristics of crowdsourced georeferencing to see where it fits within the matrix; doing so will lead to some interesting implications and suggestions for how this kind of work can be carried out in the future.

The matrix comprises quadrants (or “frames”) set on axes that represent two fundamental spectra along which any crowdsourcing effort can sit. The first axis (left to right) speaks to the participants themselves—on one end we have “Community” where participants interact directly with each other and form relationships through the project, and on the other end we have “Crowd” where participants have little or no incentive (or ability) to communicate and get to know each other. The second axis (top to bottom) speaks to the nature of a crowdsourcing effort’s internal structure and management. For this gradient, Eveleigh employs Burns and Stalker’s distinction between “organic” and “mechanistic” organization (Burns and Stalker

Figure 2.1. A user participation matrix for participatory archives (Eveleigh 2014).
1961). As she explains “…this spectrum pertains to the structural coordination of online participation, whether focused upon specific goals and objectives (mechanistic) or taking a more flexible and open-ended approach (organic)” (217). Thus we have a frame in the upper left called Collaborative Communities, in which we can place participatory archives projects that have an organic structure and a strong focus on community. A perfect example of this would be the National Archives’ crowdsourcing platform called “History Hub,” which defines itself as

…a place to ask questions, share information, work together, and find people based on their experience and interests. Experts from the National Archives as well as other experts, history enthusiasts, and citizen archivists are available to help with your research. (Osborn 2015)

The emphasis on direct interaction between participants and professional archivists and the free-from, message-board structure that are core components of this project can be easily contrasted with an example of the opposite corner of the matrix, which Eveleigh labels Transcription Machines. Transcription Machine projects have a highly regimented structure designed to control participant input, and do not encourage nor rely on the creation of a user community. This paradigm may be used to generate new content—metadata about collection items, or literal transcriptions of historic documents—by the contribution of many participants “at large.” By the People, a transcription platform built by the Library of Congress, is a perfect example of a Transcription Machine; without even creating a user account one may start transcribing by going to the website and beginning with the first page that is automatically provided—truly a “crowd”-driven endeavor (Van Hyning et al. 2022).

The third category Outreach & Engagements is in the mechanistic (not organic) and community (not crowd) quadrant of the framework, and crowdsourcing efforts of this sort tend to

---

use social media campaigns (often tied with existing web platforms such as Facebook, Flickr, or Twitter) to achieve a specific mechanistic goal. Certain transcription projects may fit this description, as discussion forums can be utilized to allow participants to discuss their work, and help each other with difficult passages. One distinct characteristic of the Outreach & Engagement strategy, and what clearly sets it apart from the free-flowing and open-ended Collaborative Community mindset, is an emphasis on predetermined goals, time-frames, and tasks. A project may seek to geo-tag all of the historic photos in a specific collection, or transcribe the papers of an important historical figure, all based on a predetermined set of content and the larger objectives of the archival institution itself.

Finally, we turn now to the fourth quadrant, and the one that may best inform a new version of collaborative georeferencing, the *Archival Commons*. This quadrant, in the top right square of Eveleigh’s matrix, is an organic (not mechanistic) and crowd- (not community-) focused vision for how the public may interact with archives. For her interpretation of the Archival Commons she draws heavily on Anderson and Allen’s article “Envisioning the Archival Commons” (Anderson and Allen 2009). The vision they lay out is detailed and extensive, but at its core is the simple idea that all users—members of the public, researchers, and professionals alike—have an equal opportunity to access and create narratives within the archive, and in this way the archive itself becomes “the commons.” They reference Giddens’ Structuration Theory which holds that structures are recursively restructured as individuals interact with them, thereby acknowledging agency of the individual to reform the environment with which they interact (Giddens 1984). As traditional archives are deeply and rigorously structured by their maintainers,
the archival commons is a place where individuals can break down and reinterpret that structure for themselves.

A core feature of their vision is that participants are able to link from one archive to another, taking advantage of existing Web standards. Users will interact with the commons not only to take in information that has been provided to them, but to create new information, perhaps by curating a new linked collection of their own, or building a new public narrative around some existing object. Given this emphasis on personal interaction, it may seem strange that the archival commons fits on the crowd (not community) end of the spectrum. Eveleigh justifies this by pointing out that with such a “web of connectivity” comes a “sense of infinite scale” (219). In other words, it is not the intention of the commons to create a bounded community as can be seen in other participatory archive initiatives described above, rather, it may be said that the goal is to create a space within which boundless communities can emerge.

Similar to “Envisioning the Archival Commons,” the article “New Uses for Old Records: A Rhizomatic Approach to Archival Access” by Duff and Haskell lays out a vision for how “rhizomatic” connections—essentially “horizontal” rather than “vertical”—between archival content and participants can build a more open and inclusive space for user engagement (Duff and Haskell 2015). They also lean on web linkage technology and social media as a means of creating non-hierarchical (i.e. from user to user, not from archivist to user) linkages and narrative creation. Thus their interpretation of “the rhizome” favors the organic (not mechanistic) and community (not crowd) side of Eveleigh’s framework. In consideration of the many participatory archive examples they give, it is clear that a dedicated community must be the basis of any effort that emphasizes the rhizomal- and open-archive-based approach for which they advocate. We
will keep this tension between crowd and community in mind as we envision the extent to which a collaborative georeferencing effort can embody principles of the archival commons.

2.1.3. Bringing the Archival Commons to Web Georeferencing

Without a doubt, one of the most daunting aspects of the archival commons is its scope. Anderson and Allen push for a “sea change in how users engage with the increasing quantities of digital objects” (Anderson and Allen 2009, 388) and Eveleigh describes their vision as one of a “global, interactive society” (Eveleigh 2014, 219). Beyond the proposed scale, or perhaps because of it, the question of who will actually populate platforms such as these is also daunting. As scholars have noted, “On the one hand, the attitude of ‘if we build it, they will come’ of some early experiments in online participation now seems comprehensively defeated” (Benoit III and Eveleigh 2019a, 216), hearkening back to the earlier and oft-cited article “Archives 2.0: If We Build It, Will They Come?” (Palmer 2009). Thus, extra attention must be paid to the consideration of who would be interested in using the sort of georeferencing application that this project envisions.

Along these lines, another core concept within the Archival Commons from which this project can take inspiration is the “Reputation of the Agents.” The authors encourage some sort of user reputation system that tracks how participants have interacted with the archive over time. To make this meaningful, the archive must be 1) long-lasting enough to allow reputation to accumulate 2) set up to record the instances in which users have interacted with materials 3) make this information available to viewers (Anderson and Allen 2009, 396). Along these lines, they also recommend easy visualization of how often users have interacted with materials. In the context of georeferencing, tracking users’ interaction with a given map or set of maps could go a
long way toward drawing people in to deeper exploration, and would also demonstrate that their participation in the materials—not only georeferencing them but simply viewing them—is valued and encouraged.

Similarly, a transparent narrative should accompany the creation of any additive information related to the materials themselves. This is exceptionally important in map georeferencing, which by its nature transforms and alters the original source material. If a participant does a poor job of georeferencing a map, it can be very distorted. Of course, providing linkages directly back to the source archive is helpful in this way—anyone can go back to look at the original if they have doubts about the transformed content. Beyond this, however, it should also be easy for users to make corrections on maps in the system that have been georeferenced by others. Further, the idea of 1) tagging each control point with the user who created it, and 2) providing the ability for commentary to be made by any user on any given control point, could further deepen this transparent narrative.

Taken together, these characteristics still do not match up squarely with the idea of the archival commons, and one reason is the underlying reliance on a technical georeferencing process, and a distinct and defined set of historic maps. However, for anything to truly grow organically, it must start small. Indeed, there is a certain irony in the very idea of “designing” a commons. Wisely, Anderson and Allen place an emphasis on practicality near the end of their article, pointing out that heritage organizations tend to lack funds and technical expertise. They call for a creation of layered networks which can be used to systematically capture the collective knowledge of those who are willing to share, and that these networks are created in a modular and maintainable fashion by organizations that have the technical knowledge and resources
available (399). It seems that a collaborative georeferencing platform can fit this description well. In terms of Eveleigh’s user participation matrix, the platform could sit between the Archival Commons and Transcription Machine frames; with a clear emphasis on accommodating an open and networked crowd, it is still grounded in a specific technical task. There is no reason of course, that Outreach & Engagement campaigns could not be built on top of this platform, but ideally the interconnected ethos of the Archival Commons should sustain public engagement.

2.2. Software Approach

2.2.1. Existing Web-based Georeferencing Platforms

Many institutions have already implemented web platforms to facilitate crowdsourced georeferencing, and though the details differ from one platform to the next, the basic procedure tends to be the same. Generally speaking, any member of the public is welcome to make a free account and begin georeferencing historical maps immediately. Once signed in, the user can see a list of un-georeferenced map scans and choose one to work with. Using a two-paneled interface, users connect points on the historical map with corresponding locations on a modern, interactive basemap, and the system warps the historical map to fit these points. The newly georeferenced map can now be viewed as an overlay in the main web map interface, and the user can move on to georeference another one.

The two most commonly implemented platforms are Map Warper, an open source application created by Tim Waters in the U.K., and Georeferencer, a commercial, proprietary platform from the Swiss company Klokan Technologies GmbH (Petr Přidal, CEO). While each of these platforms began as smaller pilot projects, they have both developed into platforms that
many different institutions have used to enhance their own map collections. Early Map Warper implementers include Harvard and Stanford Universities, Leiden Archives, and the New York Public Library (NYPL) (Waters n.d.), though none of these installations are still active. The Boston Public Library used Map Warper for many years as well, and recently the Public Record Office Victoria (PROV) in Melbourne, Australia, has implemented Map Warper (Letourneau 2019).

Because Waters published Map Warper under the open source MIT license,⁵ any institution can create its own implementation in order to focus specifically on the maps in its own collection. Alternatively, a standalone installation managed by Waters himself allows anyone to sign up and upload their own maps to georeference them.⁶ Incredibly, the website now holds over 31,000 maps, showing this to be a very desirable and popular utility for the public to have access to. Researchers have written articles about how useful Map Warper has been for their work (Aladağ 2020), and any visit to the site shows that users are uploading and georeferencing maps many times a day.

Georeferencer, on the other hand, is a commercial software offering. Many organizations currently use Georeferencer, including Leiden University and National Archives of the Netherlands (Storms 2017; van Egmond 2019), the British Library (Kowal and Přidal 2012), and the David Rumsey Map Collection (“David Rumsey Map Collection - Georeferencer” n.d.). Due to its technological origins, Georeferencer has a close affiliation with the website OldMapsOnline.com, and these systems work together as a federation of historic map collections (Southall and Pridal 2012; Cascón-Katchadourian, Ruiz-Rodríguez, and Alberich-Pascual 2018).

---

In theory, this creates a beneficial symbiotic linkage between all institutional implementations of Georeferencer and the overarching platform itself. One great benefit of this federation is that users interested in georeferencing the maps from one collection have an easy way to become involved in another collection. A potential downside, however, is that participants may find this federation confusing: if you begin the georeferencing process on the David Rumsey Map Collection site, you can very easily find yourself at a sign-in screen on georeferencer.com.

In some cases, in-house developers have built their own solutions tailored to their specific collection, and one good example of this approach is the Virtuelles Kartenforum 2.0 (“Virtual Map Forum 2.0”, and herein “VK2.0”). Created in 2014, VK2.0 was a cooperative effort between the Saxon State and University Library Dresden (SLUB Dresden) and the University of Rostock (Mendt 2014). To date, the system reports that 8,944 of 8,979 historical maps have been georeferenced. Its code is published on GitHub under the GNU General Public License v3.0 and therefore is reusable by any other institution; however, unlike the two more commonly used platforms described below, VK2.0 was purpose-built for the SLUB collection and seemingly not with the goal of being implemented by other institutions. Interestingly, though, it is tied directly to the museum’s existing content management system (CMS)—built from the open source platform TYPO3. It is an in-house solution created with a forward-thinking commitment to open data and code-sharing principles. To this author’s knowledge, no other TYPO3 users have attempted to implement VK2.0 on their own system, but presumably such a model is at least possible.

---

8 Bill and Walter (2015) provide a detailed discussion of the data quality of the georeferenced maps in VK2.0, unfortunately, in German.
Considering these platforms in the context of the archival commons, in many ways they fit well into the Engagement & Outreach category, though instead of relying on a social media platform to foster user engagement, the online georeferencing application itself serves as something of a communal space. For example, in Map Warper, users are able to comment on each map within the portion of the app where the georeferencing process occurs. However, these comments are somewhat hidden, and the number of comments per map is not shown in any of the search interfaces, making it difficult to explore the collection in a “what maps are people talking about?” manner. Neither the VK2.0 nor Georeferencer (and its many implementations, from the David Rumsey Map collection to the British Library) support user comments on maps, though they do have a system for ranking participants based on points earned by performing map georeferencing tasks (adding control points to a new map, or revising an existing set of control points, for example). These two systems also emphasize a progress report—how many maps have been georeferenced of the total online collection. And the numbers are impressive all around—the VK2.0, British Library, and David Rumsey Map Collection report 99.6, 93 and 74 percent georeferenced maps out of 8,511, 56,176, and 60,660, respectively. Map Warper on the other hand, does not prominently feature a contributor ranking or overall progress bar, and its standalone implementation at mapwarper.net is a completely open-ended, user-driven system—quite far along the “crowd” spectrum of Eveleigh’s framework.

One aspect that is common to all platforms, of course, is that they facilitate the process of georeferencing with a set of steps that even relatively non-technical users can follow. There are certain differences between the platforms—Georeferencer has some extra visual tools to ease the process, for example—but in the end, on any platform there is a fair bit of panning, zooming, and
clicking. Therein we can see the clear connection to the “mechanistic” aspect of the Outreach & Engagement paradigm. One may suppose that this amount of complexity and technicality would discourage participants, but Hedges and Dunn have made the interesting observation that users actually seemed to enjoy and be drawn in by the process. They write: “[T]hose most engaged with georeferencing as a task were far more heavily engaged with the methodological narrative of the task than the content” (Hedges and Dunn 2018, 143). As we re-imagine where collaborative georeferencing can fit in the User Participation Framework, we should have no need to shy away from the mechanistic side of the spectrum.

2.2.2. Prioritizing Open Source Software

Considering that GLAM institutions are largely built on principals of open access and public engagement, it is unsurprising that there is a well-documented preference for open source software (Huvila 2008; Fleet and Přidal 2012; Prescott 2020). Also, slim budgets and long-term funding concerns tend to push institutions away from costly licensing agreements whenever possible. Of the three crowdsourced georeferencing platforms described above, the VK2.0 and Map Warper are both fully open source projects, while Georeferencer is a proprietary platform. Interestingly, though, Georeferencer does have open source origins, as can be found in Petr Přidal and Petr Žabička’s article from 2008. The piece, published in e-Perimetron, sets forth the technological strategy behind the collaborative georeferencing backend of OldMapsOnline.com, a new project sponsored by the Czech Ministry of Culture. Therein, they conclude by writing “The online georeference tool and rectification server described above will be developed and the results will be published as an open-source project” (Přidal and Žabička 2008, 20). In reality, however, the technology behind OldMapsOnline became a key part of Georeferencer (Fleet,
Kowal, and Přidal 2012) and while some pieces of the code base are still publicly accessible,\textsuperscript{10} Georeferencer itself, as a platform that facilitates crowdsourced georeferencing, is not. The fact that Georeferencer is proprietary software came to be one of the reasons that the National Library of Scotland, one of the very first and most supportive adopters of Georeferencer (Fleet, Kowal, and Přidal 2012; Fleet and Přidal 2012) eventually discontinued its use of the platform in 2017. Specifically, staff cited “the costs of the Georeferencer application, its external hosting and control, and its proprietary nature…” (“NLS Map Georeferencer (Note)” 2017).

Of course, the lesson to take from this is clear: remain dedicated to developing software under an open source license. It is important to note here that open source software does not mean “free of cost,” only free of licensing fees. A web application must be hosted on a server, which always costs money to run, and they will likely need someone to install and maintain it. However, if open source software is good enough, and the market is wide enough, tech companies will pick it up and sell installation services or extended support contracts.

As for Map Warper, it is unfortunately based on some out-of-date technological components, and does not seem to have enough active developers to upgrade them. For one example, it uses version 2 of the JavaScript mapping library OpenLayers, which was released in 2013. OpenLayers is a well-managed, feature-rich library with frequent release, and the current release number is 6.\textsuperscript{11} The Boston Public Library cited this continued reliance on old components in its announcement that it would be deprecating its Map Warper implementation.\textsuperscript{12} One may

\textsuperscript{10} Klokantech GDAL Georeferencer, https://github.com/klokantech/gdal_georeferencer, last accessed July 7, 2022. This tool allows programmers to interact with the Georeferencer API. It is available on Github but not published under any specific license.


\textsuperscript{12} LMEC Map Warper Note, https://cartinal.leventhalmap.org/guides/lmec-dc-geo.html#accessing-maps-previously-georeferenced-in-map-warper, last accessed July 7, 2022. The note reads in full: “The Map Warper integration which was set up in our Digital Collections site is unfortunately no longer being supported. The source Map Warper software this was based on, and the crowdsourced add-on to the original software requires
conclude that Map Warper would benefit greatly from a well-placed and substantial investment into updating the core code base—ideally, such an investment could be pushed out to all other operational Map Warper installations like those at the NYPL and PROV, bringing them all up-to-date at once. In reality however, open source software upgrade pipelines can be cumbersome, so much more research would be needed before one could count on such a “ripple effect.”

In this regard, Georeferencer may actually have a strong advantage. Fleet and Přidal point out that the federated nature of the infrastructure, where all institutions are closely tied to a central, commercial platform, allows for a development pattern “whereby individual libraries pay for development work, but the results of this can then be shared in improved functionality for all partners” (Fleet and Přidal 2012, 246). Again though, it bears repeating that this commercial nature and its associated cost was cited in the discontinuation of Georeferencer by the National Library of Scotland, Fleet’s own institution.

One way that open source projects prepare for the kind of “technical debt” accumulation that Map Warper is suffering from is a further dedication to weaving together existing, mature open source projects instead of building a large new code base. In discussing how to best approach the design of community archives on the web, Prescott (2020) says they “need to make common cause with the approach taken by digital activists,” like the Mozilla Foundation. A healthy internet is one that is decentralized and does not rely on commercial platforms, and community archives should do their part to contribute to this vision (Prescott 2020, 262). Due to an abundance of open source software and very widely adopted open standards—such as the suite of data specifications developed and maintained by the Open Geospatial Consortium

---

geospatial projects on the web are very well-positioned to fully embrace a decentralized architecture. The design of a new georeferencing platform in this project will embrace these recommendations.

2.2.3. Choosing GeoNode

While Map Warper and Georeferencer are stand-alone platforms, this project took inspiration from the VK2.0 approach by beginning with an existing content management system (CMS) and adding georeferencing capabilities to it. After consideration, the geospatial CMS GeoNode stood out as the best option from which to build. GeoNode is an open source platform that is implemented by organizations and governments around the world to create open data portals for data management and dataset sharing (Corti et al. 2019). For example, one installation of GeoNode is the AlpConv Atlas, a geodata portal dedicated to promoting “research, cooperation, and monitoring activities in the Alpine Region” (Della Chiesa et al. 2021, 1). In this installation, geospatial datasets related to environmental management are stored, viewed, and combined into web mapping applications in order to create a collaborative data environment.

Technologically, GeoNode is built from a number of different open source components such as Django, PostgreSQL/PostGIS, and Geoserver.

As a CMS, GeoNode provides management capabilities for three different types of resources. Documents are uploaded files, like PDFs or digital photographs, while Layers are spatial datasets—vector formats like shapefiles, raster formats like GeoTIFFs, or remote web services (cascaded through the system) are all stored as Layers. The third resource type, Maps, are user-authored combinations of layers presented in a customized web map viewer. (Because

“map” is such a general term, especially in a project were everything seems to be some version of a map, throughout this paper and within the final software itself the GeoNode resource type Map was changed to Web Map, for clarity. Also, throughout this document, “Document,” “Layer,” and “Web Map,” will be capitalized when referring to these resources as GeoNode entities.) Conceptually, applying these GeoNode constructs to the overall georeferencing operation is very straight-forward: scanned historical maps can be uploaded as Documents and then georeferenced, creating Layers. Once a few Layers have been created, a Web Map could be authored, placing all of those Layers into a single, interactive interface. In other words, the addition of georeferencing capabilities should fit gracefully within the paradigm of the existing platform.

GeoNode provides a robust user account and permissions management system, and some of its other features are commonly seen in social networks: users may send each other messages, and they can comment on, rate, share, or “favorite” individual datasets. View counts are recorded for Documents, Layers, and Web Maps, and a Recent Activity page shows the latest resource creation actions by users. While presumably not created with Anderson and Allen’s archival commons in mind, these aspects of GeoNode do lend themselves to the idea of “Reputation of the Agents,” and, it was believed at the outset of the project, could perhaps be expanded as needed.

Another attractive aspect of GeoNode was its integration with GeoServer and the data exchange formats that this integration provides. Running behind the CMS, GeoServer creates geospatial endpoints for Layers that comply with a number of common OGC standards, the most desirable in this case being Web Map Service (WMS) and Web Map Tile Service (WMTS).
Practically speaking, this means that once a Document has been georeferenced, the resulting Layer can be used directly by any GIS user with a connection to the internet. Desktop software like ArcGIS Pro and QGIS used by researchers and GIS professionals, or external web mapping applications like those created by the City of New Orleans to serve its citizens—virtually all clients can consume WMS. In other words, through its integration with Geoserver, GeoNode provides instant publication of spatial assets through well-established protocols.

Finally, as open source software, GeoNode has an active developer community behind it, and this community is accessible through a mailing list and a public chat room.\(^{15}\) Having a community to ask questions of throughout the project did prove to be exceptionally helpful. However, GeoNode is not commonly used in the US, nor is it often implemented for cultural heritage use. This disconnect was not seen as prohibitive to the project, rather as an opportunity worth exploring. Perhaps added georeferencing capabilities could contribute to increasing GeoNode’s adoption?

### 2.3. Designing a Pilot Project

#### 2.3.1. Goals

Of course, the objective of this research was not only to create a new web georeferencing platform, but also to test that platform out through a limited public release. The most effective way to structure such a release is to focus on a specific set of historical maps and reach out not only to people with a general interest in GIS and programming, but also to those with potential interest in that set of maps and/or local history. Realistically, of course, this pilot project can only serve as a small-scale application of the archival commons ideas expressed above—server space

---

\(^{15}\) GeoNode mailing lists and chat rooms, [https://geonode.org/communication/](https://geonode.org/communication/), last accessed July 7, 2022.
to host the platform is not free, and, more importantly, the long-term, open-ended nature of the archival commons is impossible to properly foster in a short, bounded time period. Nevertheless, even limited testing of the entire concept would reveal important new ideas, and give direction to future work.

Thus, the main objectives of the pilot project (and some thoughts on how to probe them) are listed below.

1. General software evaluation
   Focus on performance, usability, and general conceptual successes and failures.

2. Track and understand user engagement
   How do people use the platform? Why do they participate in georeferencing (those that do)? Can this system efficiently facilitate crowdsourcing?

3. Generate and publish novel datasets
   Even though this pilot project is a short-term, limited participation endeavor, in the end, new georeferenced historical maps will be produced and a long-term public access plan should be put in place for their stewardship. One reason that this is a reasonable goal is that there are numerous geospatial standards and data formats through which georeferenced maps can be published—when it comes down to it, they are not complex. These files should not be lost and the strategy for their long-term storage should be considered a part of the entire process, not an afterthought.

2.3.2. Using Sanborn Maps

When considering which collection of maps would be suitable for a new effort at crowdsourced georeferencing, Sanborn fire insurance maps are an exceptionally good option.
Thematically, they provide content at a scale unparalleled by any other historical (American) map series, and their usefulness for historical research (demographic, environmental, and urban) has been acknowledged and praised many times over (Hayward 1973; Colten 1993; Oswald 1997; Berry 2003; Kodama 2013; Lafreniere and Gilliland 2015; Piekielek 2017; Lederle 2017). Even well before their publication ceased in the 1960s, the Sanborn Map Company itself had begun to acknowledge that “In recent years Sanborn Maps have been usefully employed in many fields in addition to the fire insurance industry” (Description and Utilization of the Sanborn Map 1953). Nowadays they are prominently used for environmental impact assessments because they depict old industrial complexes and fuel storage locations with great detail, and are also foundational tools for genealogical and architectural research.

Sanborn maps are an especially good fit due also to their ubiquity across the nation, and, correspondingly, their relevance for a very large audience. The entire catalog of the company’s offerings covered 7,500 cities and towns in 1912, 11,000 in 1924, and 13,000 in 1939 (Ristow 1968, 206, 211), making the scope of potentially interested stakeholders unexpectedly large and unique. Disregarding the handful of large cities that have already digitized, georeferenced, and published their own Sanborn maps, this collection provides the opportunity for a meaningful connection to almost 13,000 smaller communities—not only their residents, but also their local historical societies. Where the LOC has funded the heavy-lifting of cataloging, digitizing, and uploading hundreds of thousands of map sheets, georeferencing the content can create even more direct connection to the communities that will be most likely to engage with it. This allows the platform to engage with the crowd and community spectrum at both ends: On the one end we have a crowd of participants that can jump in and jump out at any time, contributing, perhaps as
Dunn and Hedges observed, simply because they enjoy the technological activity. On the other end, the platform could provide a space for 13,000 small digital communities to emerge, facilitating close interaction between their members based on shared geography and local history.

Another group that uses Sanborn maps very heavily consists of researchers and academics. As a primary source, the maps are incredibly useful, but researchers typically need to acquire and georeference them on their own, using desktop GIS software. What if people were able to turn to this new “archive” instead? They would be able to georeference the maps in a web browser and then use it directly in desktop software if they desired. However, if someone had already georeferenced the maps the researcher is interested in, they would benefit from the work of others before them. In this way, the system can be truly open and generative, where the usefulness of the platform only increases over time.

Therefore, the platform could exist not as an “archive” per se, but as a “meta-archive,” its contents entirely derivative (at least initially) of the existing material available through the LOC, and accessible via its API. By adding a map and georeferencing it, a user is not only creating a more geospatially useful digital object which they can download or view directly in the browser (or perhaps put directly into a web map on a city government’s own website) but is also creating a link directly back to the item’s official record in its archive of origin. This model comes with many advantages, all of which have roots in the philosophy behind the archival commons. The first most obvious one is the requirement that this platform utilize existing web APIs and spatial data transfer standards. This reliance on connectivity to outside sources is a clear priority in the archival commons, as it leaves behind the idea of a monolithic authoritative archive in favor of a distributed and more flexible one. Even though this proposed georeferencing platform would
initially prioritize a built-in linkage with a single map collection at a single archive, the design should be such that extension to other collections and later other archives will be as easy as possible.

2.3.3. Content Scope

For the pilot project, instead of focusing on the entire Sanborn map collection (which consists of over half a million individual maps!) a comparatively small subset was chosen in order to thematically focus on the state of Louisiana. The following parameters were set to limit the overall number of volumes that users could georeference:

- Include the earliest edition for every Louisiana community, regardless of date.
- Include any editions published through 1910 for all communities outside of New Orleans.
- Include only the earliest full coverage of New Orleans (in four volumes, published between 1885 and 1893).

These parameters were meant to balance appeal to different user groups—researchers wanting to see the oldest records, citizens across the state looking to learn about their hometown—while also keeping in mind some basic limitations of server space and management. Once applied, these criteria produce 267 volumes covering 138 communities, with a combined sheet count of about 1,600. Rough estimates showed that if every one of these documents did get georeferenced, the disk space needed would be about 5gb for all of the original JPEG downloads from the LOC, plus up to 100gb for the resulting GeoTIFFs.\(^\text{16}\) These are large but entirely manageable disk sizes (for the time being).

\(^{16}\) In reality, the actual georeferencing process ended up creating three copies of each GeoTIFF per layer. One of these copies was completely unnecessary and ultimately removed, the other copy was a by-product of how GeoNode managed the content, which is an important note to keep in mind for the future.
2.3.4. Potential Participants

Generally, the people, groups, and organizations to be contacted fell into the following categories:

- Social media groups related to Louisiana history, software, or GIS
- Local professional groups related to Louisiana history, software, or GIS
- Historical societies and other heritage organizations around the state
- Individuals with demonstrated interest in the subject material

The hope in reaching out to a wide range of potential participants was to expose different ways of engaging with the maps, the most obvious dichotomy being between those who just want to look at them versus those who wanted, and had the capability, to participate in the crowdsourced georeferencing aspect of the system.

2.3.5. User Survey

A survey was to be carried out through a short questionnaire, and its distribution planned for the end of the participation period. It was open to participants and non-participants alike. The questionnaire consisted of 15 questions, split into two parts. The first section contained basic questions about the participant, and the second focused on their interest in the project. The intention was to create a quick, simple, questionnaire that would provide enough information about the survey-taker to understand who they are and what interested them about the site. The full questionnaire can be found in Appendix B.
2.3.6. IRB Approval

Due to the public nature of the study, the fact that users would be signing up on the platform, and the intended distribution of the survey questionnaire, an application was submitted to the LSU Internal Review Board (IRB) in early January of 2022. On January 23, the IRB deemed the project exempt from IRB review and granted approval. A copy of the determination letter and the informed consent form that participants had to agree to before participating can be found in Appendix A.
Chapter 3. Using Sanborn Maps

3.1. History of the Sanborn Map Company

While nowadays the name “Sanborn” is practically synonymous with the entire genre of American fire insurance maps, the Sanborn Map Company was not the first to publish them. In fact, the book *Fire Insurance Maps – Their History and Application* by Diane Oswald describes the early industry as competitive and diverse (Oswald 1997). The War of 1812 caused a widespread economic break with British businesses, creating a void in the American economy that was quickly filled by new commercial entrepreneurs. Additionally, urbanization accelerated greatly in the 1820s, contributing to an increase in large fires in dense city environments. Add to this the arrival of the lithograph, which allowed maps to be reproduced extremely cheaply and quickly, and the stage was set by the 1850s for the fire insurance map industry to take off (Oswald 1997, 13–14).

It was at this time that George T. Hope recognized how insufficient the tabular register of insured properties that his insurance company used was for balancing risk; as fire spreads through adjacent buildings, a map showing the geographical dispersion of your risk is much more effective than a simple list. In 1850, Hope engaged William Perris to help make the first comprehensive map of each building in New York City, and brought together a committee of other insurance associates to oversee the project. Even, or perhaps especially, at this early stage standardization of symbols was a high priority. During the first committee meeting they determined that the new maps “should show the nature of the different materials used in the construction of all buildings by a system of colors which they named, and they also designated a few simple signs for other purposes” (Getty 1910, 20). Relaying this same passage in 1968, LOC
Chief of the Map Division Walter Ristow observed that the format and symbols decided upon by this committee “set standards for fire insurance maps that persisted, with few modifications, for more than a century” (Ristow 1968, 198).

While Perris and his later business partners continued to create maps of New York City and cities around it for decades to come, other companies popped up in different major American cities. Ernest Hexamer—who had gained experience while working for Perris in New York from 1852 to 1855—began his own company in Philadelphia, which eventually extended its focus to many other cities in the region. In St. Louis, Alphonso Whipple published a fire insurance map in 1869, and soon after created the Whipple Agency, which continued to publish for three more decades. Elsewhere in the Midwest, the Rascher Map Company was founded in 1885, creating maps of Chicago and many other Midwestern cities, and at the same time the Dakin Publishing Company mapped cities around California. By 1915, all of these companies—Perris, Hexamer, Whipple, Rascher, and Dakin (and a few more)—had merged with or been purchased by the Sanborn Map Company. Indeed, as Ristow points out, “the story of fire insurance maps in the 20th century is almost exclusively the story of the Sanborn Map Company” (Ristow 1968, 204).

In a handbook published in 1953, the company similarly boasts “...it is obvious that the entire history of insurance map making in America is included in the organization of the Sanborn Map Company” (Description and Utilization of the Sanborn Map 1953, 3).

Originally hired by Aetna Insurance to make maps of several cities in Tennessee, Daniel A. Sanborn first published his own fire insurance atlas in 1867, of Boston, and directly afterward he established the D. A. Sanborn National Insurance Diagram Bureau (Description and Utilization of the Sanborn Map 1953, 3). A clear foundation of Sanborn’s early success was
prolific surveying (he had completed 50 towns and cities in his first year) complemented by a nearly immediate copyright (in 1868) of his own map symbol scheme—which certainly owed a great deal to the standards set a decade earlier by Hope’s committee. This focus on standardization helped Sanborn gain the attention of many important figures in the insurance industry, an especially important feat as the creation of large-scale map atlases was always funded on a subscription basis. For example, the very first Perris maps of New York City were only created after 25 companies pledged $150 (Oswald 1997, 15). Another early Sanborn decision enhanced this funding model: in 1876 the newly renamed Sanborn Map Company, now with over 1,000 mapped cities in its catalog, began issuing paste-on corrections to their subscribers (Keller 1994). When an insurance company subscribed to a given city’s fire insurance map, they would first get the entire volume—a hefty, large-scale atlas, technically distributed on a lease agreement—and then receive small paper correction slips by mail to paste onto their maps, thereby keeping them up-to-date.

By the early 20th century, the Sanborn Map Company was the dominant force in fire insurance mapping, expanding its reach across the entire nation. The real engine of production for Sanborn was its fleet of field surveyors, called “striders,” who walked every street to record details of the built environment. Often this work was seen as more of a worldly apprenticeship than a job—as one strider put it, “when a young man decides to take up fire insurance surveying he is entering a field that offers more opportunities in self-education than any other profession known to the writer” (Oswald 1997, 36). Daniel Beard, a founder of the Boy Scouts of America, felt that when he began to work as a surveyor for Sanborn in 1872, his “opportunity to travel came at last” (Ristow 1968, 202).17

17 For much more about Beard’s time with the Sanborn Map Company, see Swab (2019).
The middle of the 20th century brought change to the iconic, large-format and large-scale urban atlases that had been the Sanborn Map Company’s main product. In the 1950s they introduced a new, reduced-size map (11” x 13”), and Keller points out that company literature at the time began to heavily emphasize special geographic analysis and urban development consulting services outside of the company’s traditional scope (Description and Utilization of the Sanborn Map 1953; Keller 1994, 37). Soon, the atlases were discontinued altogether. But while their contemporary value waned, their historical value was just beginning to wax; the Sanborn map archive quickly became a highly valuable, and even contentious, collection of materials.

3.2. Accessing the Library of Congress Collection

3.2.1. Early Digital Archives

In 1983 Chadwyck-Healey, Inc. of Alexandria, Va. (Sir Charles Chadwyck-Healey, founder) carried out the first effort to create a film replica of the Sanborn archive (Keller 1994, 42). The company created a black and white microfilm archive of the Sanborn maps in the Library of Congress, a collection that had grown especially large in 1967 after it incorporated the US Census Bureau’s Sanborn maps. In 1989, the environmental consulting firm ERC (Charles Slutsky, president) offered to purchase Chadwick-Healey’s entire microfilm archive, resulting in a new business partnership that utilized the microfilm archive for the purpose of environmental regulatory consulting (Oswald 1997, 80). This archive was eventually bought out by the company ProQuest in 2001, which digitized all of the microfilm and delivered the content to research institutions on a subscription basis (Information Today, Inc. 2001; Arlitsch 2002).
Smaller, more regional digital applications of Sanborn Maps began in the 1990s. One very early example was the creation of the Historical Hazard GIS, a database initiated by the Geography Program of the Illinois State Museum. This project transferred content from microfilmed Sanborn maps into a GIS database, allowing state museum staff to “produce site histories that indicate what activities occurred on individual parcels of land” (Colten 1993, 48). Like many early digitization projects, and perhaps due especially to being ahead of its time, the Historical Hazard GIS ultimately fell victim to the rapidly changing landscape of digital data formats and is no longer maintained.

As web distribution of archival material began to take off in the early 2000s, academic institutions began to eye their own Sanborn maps and felt they were finally able to create something important and novel: a publicly available color digital archive. Despite the centrality of color and symbology to the philosophy of the Sanborn creators, up to this point only black and white replicas were widely distributed on microfilm. The first of these efforts occurred in Ohio, a partnership between the Ohio Library and Information Network (OhioLINK) and the Ohio Public Library Network (OPLIN) (Bauer and Roddy 2001), and the project consisted of scanning all Sanborn maps for Ohio cities and creating a public-facing web navigation system for their viewing and retrieval. Very soon after, a similar project was undertaken by the J. Willard Marriott Library at the University of Utah, and in discussing the project, librarian Kenning Arlitsch describes the copyright challenges that they faced from the commercial world of Sanborn users. Ultimately, they only digitized content that was published before 1922, which was guaranteed to be in the public domain (Arlitsch 2002).
3.2.2. Copyright and Access

Copyright is an especially relevant topic for discussions of Sanborn maps, as the threat of infringement lawsuits has often overshadowed mass-digitization efforts like those described above. These threats come from the company EDR, an environmental research and consulting firm that purchased the entire original Sanborn archive (and its associated trademarks and copyrights) in 1996 (PE Staff 1996). With this investment, EDR was not only able to sell access to higher quality content than what was available through other vendors on microfilm, but could also tout themselves as the exclusive commercial providers of Sanborn maps. In 2018, EDR was acquired by Silver Lake and Battery Ventures, leading to the creation of the umbrella company LightBox (“LightBox Announces Acquisitions of Real Capital Markets and Digital Map Products” 2019). Slowly, it seems, the environmental assessment software products that EDR created with Sanborn maps are merging with LightBox’s real estate investment tools, though EDR currently still exists as a brand of LightBox.

In April of 2019, EDR filed a lawsuit claiming copyright infringement against another environmental services firm, ERIS. ERIS’s source of Sanborn maps actually goes back to the Chadwyck-Healey microfilm archive, a copy of which they purchased from ProQuest in 2011. EDR claims that ERIS has been selling access to maps that are still protected by copyright (“EDR Protects Its Sanborn Map® Collection” n.d.); ERIS has responded by pointing out that many, many of the copyrights EDR claims to own have actually lapsed and never been renewed. Three years later, the case is still mired in the discovery phase, while each side carefully searches through to find which specific Sanborn volumes do have current copyrights.

---

18 This sale is confirmed in court documents from the ongoing case *The Sanborn Library LLC v. ERIS Information Corporation et al.*, specifically in the “Answer & Affirmative Defenses to First Amended Complaint” page 8, filed June 17, 2019, but I have not found other sources to reference for this fact.
Ironically, where EDR must use legal action to cling to its commercial use of the Sanborn archive, copyright laws are directly responsible for the existence of the largest freely accessible online archive of Sanborn maps in the public domain, which is what brings us to the Library of Congress (Figure 3.1). Over the years, the LOC collection grew as it received a mandatory deposit of each new copyrighted work, and was further bolstered in 1967 when the US Census Bureau transferred in its entire Sanborn map holdings (Keller 1994, 42). The LOC has always been interested in making their Sanborn materials more publicly available, even going back to the original creation of the Chadwyck-Healey microfilm. In 1997 they partnered with EDR—at the time proud new owners of the original Sanborn Map Company archive—to collaborate on a

![Figure 3.1. Origins of today’s digital Sanborn map collections.](image-url)
massive digitization project of the two organizations’ combined collections, but this agreement ultimately fell through (Library of Congress Information Bulletin 1997; Arlitsch 2002).

In 2014 the library embarked on a new Sanborn digitization project, this time selecting the Midwestern company Historical Information Gatherers (HIG) to perform the scanning work (“HIG’s Digital Map Project at the Library of Congress” n.d.). HIG specializes in providing historical material to researchers, so old fire insurance maps have been an important resource for them for a long time. With the LOC agreement, HIG scanned around 35,000 Sanborn map editions that were determined at the time, one by one, to be free of copyright restrictions. As part of the agreement, HIG added these maps to their own research collections, and the library was only permitted to incrementally release the content to the public over the course of three years, between 2017 and 2020 (Maloney 2017). Alongside the scanning work, HIG also georeferenced the key maps for each map or volume, and in 2017—the same year that the library began putting their content online—HIG released a new subscription service: “Fire Insurance Maps online” or FIMo. This platform, which includes not only Sanborn maps but those from many other publishers like Hexamer and Whipple, provides an interactive web map interface for finding historical maps which is marketed to historical researchers, as opposed to EDR and ERIS’s focus on environmental risk assessment (“FIMo - Fire Insurance Maps Online” n.d.). Who subscribes to the service? Current statistics on HIG’s website state that public libraries, academic libraries, and historical/genealogical societies make up 50, 46, and 4 percent of FIMo subscribers, respectively, but provide no absolute numbers.²

Reflecting on the legacy of Sanborn maps, the collection has had two lives. The first as the flagship product of a competitive and hugely successful commercial cartography company,  

and the second as a massive archive, used and reused, transferred and transformed by different entities for different reasons. At this point, many more institutions have followed in the steps of the OhioLINK/OPLIN and University of Utah digitization efforts, and some have even georeferenced their scanned maps into large mosaics, like the University of Wisconsin–Milwaukee’s interactive *Sanborn Maps of Milwaukee* web application\(^\text{20}\), or in the Boston Public Library’s *Atlascope*.\(^\text{21}\) Still, the LOC collection stands apart as the largest publicly accessible source of high quality, color scans of Sanborn maps; while EDR and ERIS battle in court over which maps are copyrighted, the LOC guarantees that any of their digital works are fully in the public domain. Today in 2022, this includes any maps published before 1927, but a quick search of the database also reveals that about one quarter of all maps published after 1927 (3,951 of 14,843 or 27 percent) are available online.\(^\text{22}\) Thus, centering a new georeferencing platform around this map collection will open up free access to them in new and engaging ways, but, more importantly, the size of the collection will provide plenty of room for expansion, experimentation, and exploration.

### 3.3. Georeferencing Sanborn Maps

#### 3.3.1. Structure of Sanborn Maps

Before considering what a good georeferencing workflow may look like for Sanborn maps *en masse*, we must first understand how they are constructed. To do this, we will start by looking at the sheet as an atomic unit, understanding that generally it is a compilation of sheets—

---

or even a bound volume of them—that together compose a Sanborn map or atlas. Physically, the sheets were very large (21” x 25”) and they were double-mounted and stiff, built to handle the later addition of paste-on correction slips. The cartographic content of a map or atlas is spread across one or more sheets, and in the simplest cases a sheet will cover a single contiguous area. To conserve space, however, multiple discontiguous areas were often arranged onto the same sheet, separated by thick black lines. Sometimes these parts appear as small insets within the sheet, other times the sheet is divided evenly in half, or in thirds. Figure 3.2 presents a breakdown of an example sheet that contains multiples parts, and other cartographic elements as well.

On the whole, the actual cartographic content of sheets can be placed into four different categories, listed here in ascending order of scale (“zoomed out” to “zoomed in”): graphic map of volumes, key map, congested district map, and main content. These will be described below in the observed order of their prevalence across the collection, as not every map, volume, or atlas contains content falling into all four categories.

---

23 As noted above, in 1951 a new, smaller sized atlas (11” x 13”) was introduced in an attempt to reduce costs and prices. The characteristics of these editions do not figure heavily into this project, so remain unexplored here.

24 While the consistency of Sanborn maps makes it tempting to speak in generalities, an exhaustive accounting of the entire collection was far beyond the scope of this project so some common characteristics may be overlooked here.
The main content of the map or atlas is, of course, always present and is typically drawn at the same scale across all sheets, or partial sheets. The scale is usually one inch to 50 feet (approximately six to nine city blocks per sheet), though one inch to 100 feet (approximately twelve city blocks) was used to map smaller towns with less density. (In later editions, the main content sometimes included both of these scales—the larger for dense downtown areas and the smaller for residential areas.) Clearly, then, a great benefit and often the ultimate goal of georeferencing a Sanborn map or atlas is to geographically align all of these disparate pieces, creating a seamless mosaic of the main content.

Figure 3.2. Breakdown of a Sanborn map sheet—Alexandria, La., 1900, sheet 1.
The second most commonly found map content category is the *key map*, which is drawn at a much smaller (generally unspecified) scale than the main content and depicts the entire geographic extent of the main content, labeled blocks of color linking coverage areas to specific sheet numbers. Not all Sanborn maps have a key map, but when present they are generally combined with main content on the first sheet.\(^{25}\) In large volumes, the key map may cover the entire first one or two sheets, and, infrequently, an additional key map can be found buried in later pages of the volume.\(^{26}\) One important aspect of key maps is that they always show the *bulk* of the main content’s coverage, but not necessarily its *entirety*. Often, standalone industrial and commercial areas like a lumber mill or oil refinery will mapped in the main content but fall well outside the geographic extent shown in the key map.\(^{27}\)

The smallest scale content is the *graphic map of volumes*, and where the key map shows the coverage for each sheet in a map or volume, the graphic map of volumes shows the extent of each volume in an atlas. Naturally, this category is only present in large atlases comprising many volumes.

Finally, a fourth and infrequently found category of map content is the *congested district map*.\(^{28}\) These maps are drawn at a scale that fits in between the main content and the key map (one inch to 200 feet, for example) and it provides a high level overview of all buildings in downtown business districts. Interestingly, Ristow relates that at the National Board of Fire Underwriters annual meeting in 1904, they appointed a Committee of Twenty tasked specifically

\(^{25}\) Figure 3.2 provides an example of a key map combined with main content on sheet 1.

\(^{26}\) For example, sheet 53 of New Orleans, 1885, vol. 2, contains a key map for Algiers, a community across the Mississippi river from New Orleans.

\(^{27}\) A fine example is West End amusement park and commercial area, which is mapped in New Orleans 1885 vol. 2, sheet 52, but lies miles outside of the area depicted in the key map for that volume.

\(^{28}\) See Shreveport, 1906 for the only example of a congested district map encountered through the course of this project.
with investigating congested districts and the risk they posed to cities (Ristow 1968, 205).

Cursory searches of the LOC Sanborn collection for the term “congested district”—a standard note that is attached to volumes with these maps—do suggest that this feature only began to appear in Sanborn publications after 1904.

Aside from these categories of map content there are a number of non-map elements that are standard in Sanborn maps, like a title page, a key (legend), and index (Figure 3.2). Each of these can play an important role in the overall concept of digitizing Sanborn maps, but they are not seen here as foundational to the georeferencing process.

Strangely, the most difficult hierarchical level to standardize is the most important one—can a single term be applied to the collective publication of sheets for a given city and year? For small towns and cities with a small number of sheets, these were distributed as looseleaf sets, not bound into single units. Can such a tiny edition be called an atlas? Looking at the other side of the spectrum, let us consider New Orleans. The earliest publications in the LOC collection for that city are from 1885, in volumes 1 and 2. Should we refer to this as the New Orleans 1885 atlas, comprising two volumes? Perhaps, but consider that in 1887 a third “supplemental” volume was published to extend the geographic coverage of the city, and in 1893, a fourth volume was added to push coverage even further. Accordingly, sheet numbering begins in 1885 with volume 1, and continues sequentially through 1893 volume 4. Perhaps “atlas” is still the best term, but to properly reflect the date range this set of volumes would have to be called something like the “New Orleans Atlas, 1885-1893.” That said, the Sanborn Map Company seems never to have used the term atlas, preferring instead, simply, “Sanborn Map.”
To somewhat sidestep the entire question, this project looked to the organizational structure of the LOC collection. Therein, each separate volume for New Orleans is an item, alongside each set of sheets for a small town like Abbeville in 1885. Thus, the generic term “volume” was chosen, largely to properly describe the New Orleans volumes, but with the understanding that though it is not the best term to describe a single sheet edition (e.g. Baker, 1922), it is also not the worst.

A further complicating feature is the basic concept of dating a volume at all. The LOC collection is, primarily, made up of original copies acquired through mandatory copyright deposit, meaning that they were never corrected via Sanborn’s paste-on subscription system, and their publication dates properly reflect the content in the maps. However, a secondary Sanborn series in the collection was transferred in from the US Census Bureau, and these maps were corrected with paste-on slips. This means that an edition published in 1940 may actually have been kept up-to-date through 1950 (which can be tracked in the volume’s correction record), and in the LOC database such an item would have a publication date listed as “1940 - Jul 1950” attached to it. Luckily, all volumes used in this project were uncorrected originals (i.e. not part of the US Census Bureau transfer) so comprehensively addressing this date complexity was unnecessary.

3.3.2. Workflow

With consideration to the various components and characteristics of Sanborn maps as described above, the following approach emerges for georeferencing an entire set of maps—whether for a few sheets or an entire atlas—the objective being a seamless mosaic of the main content:
1. Define the boundaries of all cartographic content across all sheets
   ○ Split multi-part sheets along divider lines as necessary
2. Georeference the graphic map of volumes (if present)
3. Georeference the key map (if present)
   ○ Use the graphic map of volumes as a reference layer to aid this process
4. Georeference each sheet or partial sheet of main content
   ○ Use the key map as a reference layer to aid this process
5. Trim the edges of overlapping sheets to create a seamless mosaic

Of course, such a basic treatment of the workflow raises many technical questions. For example, how exactly should multi-part sheets be split? If the edges of overlapping sheets must be trimmed, could this trimming be combined with the original definition of the individual parts? How can over-zealous trimming—which causes ugly gaps between layers, visually breaking the mosaic—be prevented? The georeferencing process itself raises its own questions: How many control points must be used? Is an error calculation necessary? What transformation algorithms should be applied?

Due to the technical nature of these considerations they will not be addressed directly here, but will be kept in mind and referenced later in the more detailed description of the software development process.
Chapter 4. Building the Platform

4.1. Overview of the Development Process

4.1.1. Layered Software Architecture

Once the decision to base the platform on GeoNode was made and the desire to build around the LOC Sanborn Map collection identified, a multi-layered pattern emerged for the design of the new software components. A first layer, herein referred to as the “georeference extension,” would facilitate the georeferencing process and be designed as a generic GeoNode module while a second, independent layer, herein the “LOC Insurance Maps app,” would provide structure for importing, grouping, and exploring Sanborn map volumes, in addition to holding the custom theme for the site (Figure 4.1). Altogether, the name “Louisiana Historical Map Georeferencer” (LaHMG) was chosen for the site as a whole, and it was deployed to the domain oldinsurancemaps.net. This modular approach to the code development would provide for the possibility of implementing the georeferencing extension within any other GeoNode
installation, while also integrating georeferencing capabilities into the center of this customized
LOC Insurance Maps app installation of GeoNode.29

Using a layered approach like this is common in open source web development—it allows the core code base (GeoNode in this case) to be used without modification while extra, custom functionality is layered on top. In theory, this allows for relatively easy upgrades to the underlying software when new releases are made. However, the more heavily dependent the “downstream” custom components are on the “upstream” core software, the more likely they are to need maintenance when underlying upgrades are carried out.

Putting all of these ideas together, Figure 4.2 shows the overall architecture of the platform. In the center, the GeoNode installation itself holds the imported documents from the

---

29 GeoNode is built from the web framework Django, and in the Django paradigm the “georeference extension” and “LOC Insurance Maps app” components described here are both “apps” (in the code base, they are named georeference and loc_insurancemaps, respectively). The terms used in this thesis will, hopefully, provide more clarity on the functional distinction between the two.
LOC collection, and these are imported through the “Load Volume” capability of the LOC Insurance Maps app. Once in the system as Documents, they can be processed by crowdsourcing participants through the georeference app, turning them into Layers. The Volume Summary page of the LOC Insurance Maps app aggregates and presents these layers, but software developers and researchers can also view them directly in external mapping software or combine them into Web Maps within GeoNode itself. Users, then, are really interacting with all levels of the architecture—default GeoNode pages, the branding and Sanborn structure that the LOC Insurance Maps app provides, and the various georeference extension interfaces—without necessarily being aware of the underlying layered structure.

4.1.2. Open Source License

All code created through the course of this project is published under the open GNU General Public License v3.0 and available to view, fork, or comment on in a public GitHub repository.30

4.2. The Georeference Extension

The georeference extension facilitates the actual georeferencing process, allowing users to georeference Documents in GeoNode and create spatial Layers from them. From a user’s perspective, the extension consists of three tool interfaces, as well as a new “Georeference” summary tab in the Document and Layer detail pages.

To present the extension, I will first lay out some underlying principles that guided its design, followed by a high-level summary of the process itself. A description of each tool will follow, and the section will conclude with a graphical data model and explanation.

4.2.1. Design Principles

Three underlying principles guided the structure of the workflow: 1) georeferencing as metadata creation, 2) iterative and collaborative work, and 3) independent stages.

4.2.1.1. Georeferencing as Data Creation

The best approach to georeferencing sees the process as metadata creation, not as dataset creation. In other words, it is much more important to create and store GCPs for each document, than to generate the resulting warped layer. With this in mind, the purpose of the georeference extension is not to georeference historical maps, *per se*, but to create a system in which the data needed for georeferencing—namely, GCPs—can be created and stored. Thus, the resulting geospatial layer is disposable because at any point it can be recreated by re-processing the GCPs that are attached to the original historical map scan.

4.2.1.2. Iterative and Collaborative Work

As much as possible, crowdsourced work should be performed iteratively and collaboratively. This is because there is no guarantee that the first person to georeference a document will do it well, and, more importantly, the second person to georeference that document may do it even worse. Thus, discrete work activities should be recorded as they occur, and provisions made for the “reapplication” of previous work sessions at any point in time. Along these lines, if a process fails to complete, it should be possible to re-run it with all of the same input data.

---

31 Quantitative quality control of georeferencing results did not figure heavily into how these principles were implemented. One reason for this is that due to their large scale and planimetric accuracy, quick visual checks tend to be sufficient to determine whether Sanborn maps have been correctly georeferenced. Thus, the approach here prioritized self-corrective operations, akin to Wikipedia or Open Street Map. However, this in no way precludes the later addition of quantitative GCP analysis.
4.2.1.3. Independent Stages

The entire georeferencing process for a single document may consist of multiple steps, and these should be carried out independently, potentially by different users at different times. This modular approach should facilitate as wide a variety of user experiences as possible, not only to accommodate different skill levels—some steps may be more technically challenging than others—but also different interests and personal motivation.

4.2.2. Approach

4.2.2.1. The Process Envisioned

Ultimately, three steps were designed to carry out the entire georeferencing process for a given Document. While they fit the needs of the Sanborn map georeferencing methodology outlined at the end of the previous chapter, they are also generic enough to accommodate a very wide variety of georeferencing applications were the extension to be added to a different GeoNode installation. Assuming a Document has already been loaded into the system, the steps are:

1. Preparation - *Define the area(s) to be georeferenced.*
2. Georeference - *Create ground control points.*
3. Trim (optional) - *Remove the edges of a layer if desired.*

These steps need to be completed sequentially, but they need not be completed all in one sitting nor by the same user. Preparation is a one-time step, while georeferencing and trimming are iterative and can be revisited at any time. Trimming is considered an optional step because while
it may be necessary to remove the margins of adjacent layers to avoid overlap, it is generally not needed for standalone layers.

The difference between the preparation and trim steps deserves a bit of explanation, because in some applications they are conflated within a more general operation called “masking.” Even though both steps involve adjusting the dimensions of the image in some way, they serve fundamentally different purposes. Consider that in a GCP, a specific point on the image—an x, y pixel coordinate—is linked to a geographic location—a latitude/longitude position.\textsuperscript{32} Now, if the size of the image changes after the GCP has been created, the pixel coordinate in the GCP will correspond to a different location in the altered image, rendering the linkage meaningless. This means that before any GCPs are created, the image boundary within which they will be place must be defined and guaranteed to not change later.

On the other hand, trimming is performed after the new layer has been created and does not alter the underlying file at all—the user simply defines a polygon which is used to hide any content that falls outside of it. This approach retains the ability to revert to a layer style that shows the entire original document, which may be desirable in some applications.

\subsection*{4.2.2.2. Implementing Sessions}

Combining the concepts of “georeferencing as metadata creation” and “iterative work,” it is clear that two complementary storage plans are necessary—canonical data vs. session data. This was accomplished by the implementation of database models (\texttt{PrepSession}, \texttt{GeorefSession}, and \texttt{TrimSession}) that store the input data from a user’s work on any given

\textsuperscript{32} In reality, a projected coordinate reference system is generally used for georeferencing, and this project used the \textit{WGS 84 Pseudo-Mercator} projection (EPSG 3857), which is standard in web maps. Thus, the actual geographic component of a GCP is not latitude/longitude, but a “northing” and “easting” in meters from the projection’s origin.
step. Only after the session object and its data are saved to the database is its `run()` method invoked, performing the processing step with its stored data as input. In the case of georeferencing, for example, a user will create a set of GCPs, and these are saved in a new `GeorefSession` instance. Then the session is run, which updates the canonical group of GCPs for the Document, and uses them to warp the original image and register a new Layer.

The session model provides other helpful utilities. It quickly became clear throughout the project that sometimes a session would need to be reverted outright—the input data had been corrupted, or the user had not performed the step correctly. In other instances, a session would fail while it was running, and the system administrator would need to manually “re-run” it. Additionally, sessions store two timestamps: one for their initial creation and one for the last time they were run. Due to the fact that a session is initially created whenever a user begins working on a task and then run when they have finished the task, it is generally possible to calculate the elapsed time spent by a user on each one.\(^\text{33}\)

Sessions also enable the creation of a “locking” mechanism for georeferencing steps. Once a user has entered one of the tool interfaces (described below) a new session is created and linked to the resource they are working on. When other users try to access the same interface, they will encounter a locked interface. This feature of sessions was not implemented at the beginning of the project, but it was added later to provide for the possibility of simultaneous activity by multiple users on many different documents in the same volume.

\(^{33}\) Calculating time spent in a session was not an initial design priority, but yields interesting information, as described in the following chapter. However, if a session was re-run, its user-input time can no longer be calculated because it is not reflected by the interval between initial creation and time of execution. Improving this model would be trivial.
4.2.2.3. Tracking Steps with Thesaurus Keywords

The extension keeps track of the georeferencing steps that have been performed on a given resource by taking advantage of GeoNode’s *thesaurus keyword* system. This system allows users to generate their own hierarchical collection of keywords and attach them to resources in the CMS. A new “Georeferencing Status” thesaurus contains a set of keywords that corresponds to each specific stage in the overall georeferencing process (Table 4.1).

Table 4.1. Georeferencing Status thesaurus keywords.

<table>
<thead>
<tr>
<th>Keyword Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unprepared</td>
<td>Attaching this keyword to a Document begins the process.</td>
</tr>
<tr>
<td>Preparing – in progress</td>
<td>Indicates a preparation session is in progress (or running).</td>
</tr>
<tr>
<td>Prepared</td>
<td>The Document is ready to be georeferenced.</td>
</tr>
<tr>
<td>Georeferencing – in progress</td>
<td>Indicates a georeferencing session is in progress (or running).</td>
</tr>
<tr>
<td>Georeferenced</td>
<td>The Document has been georeferenced; the Layer was made from a georeferenced document.</td>
</tr>
<tr>
<td>Trimming – in progress</td>
<td>Indicates a trimming session is in progress (or running).</td>
</tr>
<tr>
<td>Trimmed</td>
<td>The Layer has been trimmed.</td>
</tr>
</tbody>
</table>

As implemented, these keywords become a new facet in the default GeoNode search pages for Documents and Layers, allowing users to, for example, quickly find all of the documents that are ready to be georeferenced (search for “Prepared”) or see how many are currently being processed (search for “Georeferencing – in progress”).

Moreover, context processors and middleware functions are able to interrogate the keyword associated with a resource to show or hide links to the tool interfaces described below, based on the next step that should be performed for a resource. Sessions, as described above, manage these keywords as well: when a user begins a task, the appropriate “in progress” keyword is assigned, and when the session run is successfully completed, the keyword for the
next appropriate stage is assigned. Similarly, if a user uploads a new Document to GeoNode, setting its Georeferencing Status to “Unprepared” is the only action needed to enable the use of the georeferencing extension on that Document.

4.2.3. User Interfaces

The web browser tool for each step can be accessed by any user, but only registered users can use them to perform tasks. Links to the next step for a given resource were added within default GeoNode templates, in an effort to weave the extension into the existing system. A description of each tool interface will be provided here, followed by technical details on the implementation. All of the tools were built as single-page applications with the front-end framework Svelte, and OpenLayers is used for all of the mapping interfaces.

4.2.3.1. Preparation/Split

To facilitate the preparation step, the first tool presents the entire document to the user for inspection. If the Document only depicts one map, the user need only click the “No Split Needed” button, and they will be brought directly to the georeferencing interface. If the Document does need to be split, the user will draw straight “cut-lines” across the entire image, virtually “cutting” it into pieces until each map is separated into its own region (Figure 4.3). Using cut-lines instead of polygons provides a very simple and quick user experience, and guarantees that the document is split into mutually exclusive regions—no visual information from the historical map is lost. Once the user-generated divisions, shown in a preview with yellow boundaries, properly reflect the divisions on the original image, the user clicks “Split”
and the splitting process will begin running in the background (they are sent to the Document
detail page in the meantime).

The mechanics behind this interface depend on a call and response interaction between
the front-end and the back-end of the application. When the user draws or modifies a cut-line, all
lines are collected and sent to the server to generate the preview. This is done by iterating the list
of cut-lines and using the PostGIS \texttt{ST\_Split} function to compare each one to a polygon
representing the border of the original image. If the line fully intersects the polygon, the split is
performed and then the two resulting polygons are each checked against the second line in the

![Figure 4.3. The preparation process. This Document is prepared by drawing two “cut-lines”
across it, splitting it into three pieces.](image)

list, and so on. The coordinates for the final divisions are then sent back to the interface, and used
to added a preview boundaries for user inspection. This allows for any number of page divisions
to be made, as if the user were simply wielding a pair of scissors and cutting the document into
pieces.
When the split divisions have been properly defined and the user submits the cut-lines, the divisions are generated a final time and both the cut-lines and division polygons are saved to the PrepSession. Then the session run() method is invoked to process this data, splitting the image into new files on disk, generating new Documents from each one, and assigning the Prepared keyword to them. The original document is then marked as metadata_only, a GeoNode setting used to retain a resource in the system but hide it from search results. Finally, all Georeferencing Status keywords are removed from the original document, effectively removing it from the overall georeferencing process altogether. In its place, each of the individual new Documents will be georeferenced separately.

4.2.3.2. Georeference

The georeferencing interface facilitates the creation of GCPs. A dual panel design mimics most other georeferencing interfaces in existence: one panel shows the original document and the other shows a standard web map (Figure 4.4). Creating a GCP is achieved by clicking once on the original document, and then a second time on the web map. As GCPs are added, a running list of them appears in the bottom left menu, and the user can attach a note to each one if desired. Controls in the bottom right provide management of different layer display options, and allow the user to choose between two different rectification algorithms, Polynomial (1\textsuperscript{st} order only) and Thin Plate Spline.
One important feature of the interface is a live preview showing the warped original map based on the current set of GCPs. This preview appears automatically after three GCPs have been created. Providing this preview was a very high priority during the development of the interface, and Figure 4.5 shows the final strategy for implementing it. The preview is generated in the following manner: When a user begins a georeferencing session, a new virtual raster (VRT) file is created on disk alongside the original image. A VRT is a very lightweight XML file that only holds a pointer to the original image file (it stores no actual image data), but it can also hold extra metadata like a coordinate reference system and GCPs.\(^{35}\) Whenever the user creates or

\(^{35}\) Traditionally, VRTs are used to store pointers to multiple files, creating a “virtual mosaic” from multiple sources. This aspect of them was not utilized in the project, but could be in the future. To learn more about
modifies GCPs in the web interface, the updated list of GCPs is embedded into the VRT. MapServer—a lightweight spatial data server—then publishes the VRT as a WMS through Apache webserver. The web interface has been provided at the outset with the anticipated url for this WMS, so once there are enough GCPs this layer is added to the map. Though it may sound like a lot of steps, VRTs are such small files that updating them is near-instantaneous and MapServer uses the new VRT immediately, always making the latest update available in the WMS. The web map need only trigger a refresh on the layer for the interface to reflect the updated preview. To the user, the experience is quite smooth.

Once a user is satisfied with the GCPs, they can be will submitted. At this point, the `GeorefSession` is saved with the GCP data, and the same process that created the VRT preview is repeated, only now producing a Cloud Optimized GeoTIFF (COG) instead. Finally, this new GeoTIFF is supplied to a mock “upload session” in GeoNode, creating a new data store and layer

---

36 VRTs, see https://gdal.org/drivers/raster/vrt.html, last accessed July 7, 2022.
36 An extension can be added to GeoServer that allows it to read and serve VRTs, so naturally this was the first approach attempted, given that Geoserver is already an integral piece of the platform. However, while this method did seem to work at first, refreshing the source (which must happen each time a GCP is altered) was not instantaneous, causing the preview layer in the interface to retain tile artifacts. MapServer solved this issue and natively reads VRTs, no extension needed, so it was chosen for this small piece of the application.
in Geoserver. The Georeferencing Status “Georeferenced” is now set on the original Document resource, and also on the new Layer resource.

4.2.3.3. Trim

The “trim” interface allows a user to draw a mask on top of a layer in order to remove the margins of scanned maps. Doing so is not absolutely necessary, but it does allow adjacent map sheets from a historical map series to be overlapped, creating a seamless mosaic. Additionally, as shown in Figure 4.6, it can be used to remove non-cartographic content that is present on other parts of the original document. This process does not alter the file. Instead, it stores a mask polygon for the layer, generates a new GeoServer style by passing the mask polygon to the CropCoverage SLD filter, and then sets this style as the default. While a user is constructing the

![Figure 4.6. The trim interface facilitates the removal of margins and non-cartographic content.](image)
polygon, the CropCoverage SLD is mocked in on the front-end and passed to OpenLayers to preview the mask.

4.2.3.4. The Georeference Info Tab

Within the Document and Layer detail pages, which display the full set of GeoNode-related metadata about a resource, a new Georeference tab was added alongside the existing tabs —Info, Share, Ratings, Comments, and Favorite. This new tab provides a complete overview of all georeferencing actions that have been undertaken for the given Document or Layer. It is not only a central location to find the georeferencing status of a resource, but as multiple users participate by performing different tasks—at different points in time—it begins to create a narrative of public interaction with the content (Figure 4.7).
Figure 4.7. The Georeference tab shows a full history of actions performed on a resource.
4.2.4. Data Model

Figure 4.8 illustrates the data model for the georeference extension. The difference between canonical data and session data as described in section 4.2.2.2. is apparent. To explain the former, each Document is attached to a single GCPGroup object, and this group aggregates all of the GCP objects currently used for the Document. Similarly, a canonical LayerMask exists for each Layer that has been trimmed, which holds the polygon for the mask and updates each time a user modifies it. On the other hand, all session data is stored in instances of the SessionBase model, shown near the center of the diagram. The individual session types that are used to record each georeferencing task—PrepSession, GeorefSession, and TrimSession—are proxies of this base model, meaning that they inherit the same set of fields (like status, user, date_created, and date_run) but have different behavioral methods attached to them. Further, and more importantly, the SessionBase.data field stores a different JSON data structure for...
each step, fully capturing the raw user input. See Appendix C for examples of these three different data structures.

4.3. The LOC Insurance Maps App

4.3.1. Overview

While the georeference extension provides the capability to georeference individual Documents within GeoNode, further structure was needed to facilitate a connection to the LOC Sanborn Map collection and the complexity of the maps themselves. Thus, the purposes behind the LOC Insurance Maps app were as follows:

1. Manage the ingestion of Sanborn maps from the LOC online collection

2. Curate the presentation of Sanborn maps and the GeoNode resources for each “volume”

Additionally, GeoNode provides the ability to create a custom theme (colors, logo) which was stored in this app, and extra work was performed to improve the look and feel of the website for this particular implementation. For example, the default icons for different resource types were replaced with icons drawn by two separate artists found on the Noun Project.38

4.3.2. Accessing the LOC Collection

While one approach would have been to preload the entire system with all desired Sanborn maps (download files from the LOC website and then upload them to GeoNode), this did not fit the idea of an extensible, organically growing collection that is central to the concept of the archival commons. Thus, an alternative approach was taken that would require users to

---

38 The Noun Project is a repository of icons and photographs contributed by artists around the world. Icons used in the LaHMG were drawn by Alex Muravev (https://thenounproject.com/alex2900) and Olga (https://thenounproject.com/olgamur_2015). The icons were downloaded via a “NounPro” account, which permits their usage without attribution.
deliberately “load” a volume into the system, after determining they were interested in georeferencing it. This loading action, done through the “volume summary” page described in further detail below, utilized the LOC JSON API to find and download all sheets of a volume. The process of ingesting a single image (scanned map sheet) was implemented as follows:

1. Download the image in JPEG 2000 format (.jp2)
2. Convert the file to JPEG format (.jpg)
3. Create a new Document resource in GeoNode
4. Attach the JPEG to the Document
5. Set a number of GeoNode’s default attributes on the Document
6. Set the Georeferencing Status to “Unprepared”

One important aspect of step five is the incorporation of GeoNode’s “Region” attribute. Regions are predefined geographic areas that can be attached to resources, thereby providing the ability to perform spatial queries. A good bit of preprocessing was necessary ahead of time to create Region objects for this project. Queries to the LOC API generated a list of all city and town names for the maps in Louisiana, and this list was cross-referenced with a US Census Bureau “Places” cartographic boundary shapefile for 2020. For each match, a bounding box of the corresponding area was calculated and used to create a GeoNode Region object. In almost every case the Sanborn map location clearly corresponded to a feature in the Census Bureau dataset, notable exceptions being Alto, Bayou Sara, Hot Wells, and Loranger, all of which were added by hand easily enough. Sulphur Mines, a company mine site near Lake Charles, was not located at

the outset, but eventually one of the crowdsourcing participants found this community by using historical USGS topographic maps.

Parsing the item information provided by the LOC JSON API for a given Sanborn volume was fairly straightforward, with a few complications. For example, the API provides location attributes without a hierarchy, so while the terms “Louisiana,” “Orleans Parish,” and “New Orleans” may all be attached to an item, nothing specifically indicates which one of them represents the state, parish, or city name. Ultimately, cross-referencing these terms with the title of the item (typically constructed as “Sanborn Fire Insurance Map from New Orleans, Orleans Parish, Louisiana”), a list of all state names, and searching for the word “parish” was sufficient. The LOC collection also has some misspellings in it, like “Sanit James Parish,” or “Keatchie” instead of “Keachi.” A lookup file was created to handle these inconsistencies. With these challenges in mind, it became clear that while the entire system was created to be as generic as possible, some data preparation would be needed to extend the effort beyond Louisiana.

4.3.3. Managing Volumes

Perhaps the most important aspect of the LOC Insurance Maps app is the aggregation of Documents and Layers by Sanborn map volume. While the attachment of Region objects and publication dates to each Document does make it possible to use the default GeoNode search capabilities to find all Documents for a specific Sanborn edition (Region = Lake Charles, Date = 1903, for example), and the addition of the Georeferencing Status keyword “Prepared” would show all Documents from the edition that are ready to be georeferenced, this method leaves much to be desired as a comprehensive summary and accessible visualization of the
georeferencing progress that has been made across all sheets in Lake Charles, 1903. A new “volume summary” page addressed this (Figure 4.9).

In a volume summary there are two main sections: the Map Overview and the Georeferencing Overview. The latter has three collapsible subsections corresponding to different stages of the georeferencing process—Unprepared, Prepared, and Georeferenced (no separate
section was created to hold trimmed layers, given the optional nature of that step). All resources for the volume are sorted into one of these categories and represented by thumbnails. To elaborate: When a volume is first loaded, all Documents will appear in the Unprepared section, each with a link to the preparation interface. If a user prepares a Document and needs to split it into three pieces, the three new Documents will appear in the Prepared section, and the original Document will be removed from the volume summary entirely. Then, as each split piece is georeferenced it is removed from the Prepared section and the resulting Layer appears in the Georeferenced section. This sorting is determined by the Georeferencing Status keyword attached to each resource. The effect of these sections is to provide an easy-to-understand tabular progress overview, as well as contextualized links to the tool interfaces for each item.

The Map Overview section did not exist at the initial public release of the platform, but was added soon after as the need for it quickly became clear. The section contains a collapsible Preview Map that automatically aggregates all of the Layers for this volume (i.e. all items listed in the Georeferenced section) into a single interactive web map. This preview map was not part of the initial design because GeoNode provides the capability for users to create their own Web Maps from Layers in the system. The idea, then, was that users would naturally combine Layers for a whole city in one of these Web Maps. While possible, it was immediately obvious this was a cumbersome solution: First, someone would have to create the Web Map object, and then, each time a new Document was georeferenced in the volume, its Layer would have to be manually added to the Web Map. An automatically generated solution was sought and the Map Overview

---

40 A new thumbnail creation method for Documents was implemented specifically for this display. Instead of creating a consistently-sized, clipped version of the original image as GeoNode does by default, the new method shrinks the entire Document image to fit within 200 x 200 pixels. This retains the abnormal shapes of split pieces, presenting an intuitive visual summary of the volume’s content.
was the result, initially added within a week of the public release, and significantly upgraded a few weeks later.

Though a viewer of very low complexity, the effect of the Preview Map is to provide immediate visual feedback from the georeferencing process, showing users how their work fits into the mosaic of maps for the volume, building out the fabric of the historic city one piece at a time. The viewer also offered the opportunity to reflect the internal structure of a Sanborn map, i.e. the different categories of cartographic content as described in section 3.3.1. Users can designate one or more Layers as the volume’s key map, which allows the Preview Map to place these layers in their own category underneath the main content of the volume (again, see Figure 4.9). Applied in this manner, the key map provides an even clearer context for the “completeness” of the volume—the colored rectangles of the key map indicating the location of each sheet are slowly obscured by the georeferenced sheet itself.41 The key map was used advantageously in one other way: once one or more Layers had been designated as the key map for a volume, following the “georeference” link for a Prepared document from the volume summary page would add the key map as a reference layer in the georeferencing interface.

To support trimming and overlap curation—to make sure that a seamless mosaic can be created without gaps between the layers—signed in users can click the Arrange Layers button, revealing an ordered list of all layers in the main content and showing which ones appear “on top” of their neighbors. Users can adjust this list, moving one layer under or over others. This

---

41 With regard to the other categories of cartographic content, the graphic map of volumes and congested district map: The former was not present in any of the volumes used in this project, and the latter occurred only once. (In fact, when a user reported the congested district map in Shreveport 1909, it was an unknown feature to this author!) Handling these two categories could be accomplished easily enough by following the key map > main content pattern shown here.
mechanism makes it possible to predictably trim only the overlapping layers, leaving lower ones untouched to fill the gaps as needed.

4.3.4. Data Model

The data model for the LOC Insurance Maps app is not especially complex, predominately consisting of one model, Volume, which holds all of the information needed to populate a volume summary page, as well as extra information like which user initiated the load for its Sheets (Figure 4.10). The Sheet model is the main connector between Volumes and Documents, while the FullThumbnail model holds the custom thumbnail images created for this app.

![Diagram of data model](image)

Figure 4.10. Data model for the LOC Insurance Maps app.
4.4. Documentation

Comprehensive documentation was created and published via ReadTheDocs, a free documentation hosting platform commonly used for open source software projects.\footnote{LaHMG Documentation, \url{https://docs.oldinsurancemaps.net}, last accessed July 7, 2022.} This set of documentation includes background on the project and Sanborn maps, guides for how to use the georeferencing tools, a full tutorial, and extra information related to technical aspects of the site. While the georeferencing tools section focused on the purpose for and how to use each one, the tutorial used the entire volume of Alexandria, 1900 to illustrate how the process worked as a whole, providing an overview of how to go from loading the volume to generating a web map mosaic of its content.

Additionally, an effort was made within the platform itself to provide information to users on how to use its components. Each of the georeferencing tool interfaces had a help bar across the top with expandable sections containing information (see Figures 4.4 and 4.6). At the initial release of the project, the video link was not present but later an instructional video was eventually recorded and uploaded to YouTube.

4.5. Notes on the Development Process

The decision to base the platform on GeoNode was made in the fall of 2020, and local development and prototyping began around that time. Through the winter and spring of 2020/2021, potential collaboration with Dutch developer Bert Spaan was explored, though ultimately not pursued (see section 6.1.5 for further discussion on this topic).
to the project’s public repository took place in July of 2021, and since then more than 500
commits have been made, resulting in approximately 12,000 original lines of code.\textsuperscript{43}

GitHub issue tracking was used to record bugs and potential enhancements to the
platform, especially after the beginning of the public participation period in February 2022.\textsuperscript{44}
This author and one heavy user of the platform were the only ticket creators. Through the end of
the public participation period, May 2022, 77 issues had been opened, 42 of which have been
addressed and closed. Of these 42 issues, 28 were categorized as “bugs” and the remaining 14 as
“enhancements.” These numbers are not presented here to report an exact accounting of
problems found and fixes made, rather, they are meant to serve only as evidence that active
development took place after the public opening of the project.

No official “release” was tagged in the repository, leaving it in a \textit{de facto} “beta testing”
state through the course of the project. In all, the platform was in development for about 15
months, on and off around other academic and professional commitments.

\textsuperscript{43} The entire code base of the two apps has a much higher number of lines in it, but a rough estimate of 12,000
comes from an attempt to count only original material, excluding boilerplate GeoNode, Django, or Svelte code.
\textsuperscript{44} GitHub repository issue tracking, \url{https://github.com/mradamcox/loc-insurancemaps/issues}, last accessed July 7, 2022.
Chapter 5. Pilot Project Results

5.1. Overview

During the public participation period of the pilot project any member of the public was able to sign up and georeference Sanborn maps of Louisiana on the new platform. Due to a longer than expected development process, the start date of this release was pushed back, but everything was ready at the beginning of February 2022. The duration of the public period was initially planned for two months, but near the end of March an extension was approved by the LSU IRB to double the length of public participation. This extra time not only allowed for more outreach efforts and georeferencing to be completed, but also allowed users to take advantage of software improvements (and bug-fixes) that had been deployed along the way.

By the end of the participation period on May 31st, 66 users had registered to participate. All 267 volumes that had been made available at the outside had been started, meaning that all of their 1,600 sheets had been loaded into the system. Further, almost every sheet had at least been moved through the “preparation” step of the georeferencing process.\(^{45}\) Remember, preparing a sheet may split it into multiple Documents, each to be georeferenced individually, so in the end the system contained almost 2,700 Documents. Users georeferenced 1,500 of these, and 185 of the resulting Layers were trimmed.

The user questionnaire was created on Qualtrics and distributed via the email contact list one week before the end of the participation period, and a link to it was placed in a banner at the top of the project website. In the final wrap-up email announcement sent out on May 31\(^{st}\), the link was included again. This chapter will illustrate the results further, first by providing more

\(^{45}\) As will be discussed below in further detail, some entire Sanborn map sheets are filled with a text index and contain no cartographic content. This accounts for the fact that not every single sheet was prepared.
detail about who got involved and how, next with a detailed look at the georeferencing activity, and finally with a short summary of the user survey results. The following chapter will discuss the implications of these results.

5.2. Outreach Strategies

Outreach occurred through multiple channels: social media posts, e-mail blasts, paper flyers placed in public spaces (Figure 5.1), and, naturally, personal conversations with individuals who had a general interest in the subject matter. Additionally, numerous presentations were made at small conferences, for professional groups, or advertised at large for anyone to join. Generally, the people, groups, and organizations contacted fell into four categories. A description of these categories is below, followed by a chronological summary of the dates on which announcements and other outreach activities were actually carried out.

![georeference historical fire insurance maps of Louisiana](image)

**How does it work?**
- Use web browser-based tools to georeference maps from the digital Sanborn Map Collection at the Library of Congress
- Create interactive web map overlays and seamless mosaics that allow members of the public to research and explore Louisiana history

**What do the maps show?**
- Fire insurance maps contain a wealth of urban and architectural detail from the past (1885-1950)
- 1,600 map sheets covering 138 communities across Louisiana, from Abbeville to Youngsville

*This study has been approved by the LSU Institutional Review Board (IRB).*

Learn more and participate!
OLDINSURANCEMAPS.NET

Figure 5.1. Half-page flyer used for outreach.
5.2.1. Participant Categories

5.2.1.1. Social Media Groups

Social media groups were chosen that have a general focus in local history, maps, or GIS. Though all of these groups were eventually contacted (i.e. posts were made on their various public channels), only Abandoned Louisiana was included at the very beginning of the project. In fact, Reddit was not even considered during the planning process, but was added after the decision to extend the public participation period was made.

- Abandoned Louisiana
  - A state-wide Facebook group where people post pictures of and exchange information about abandoned buildings around the state

- The Spatial Community
  - A Slack group where GIS professionals around the world share information

- Nola Devs
  - A Slack group for software developers in New Orleans

- Reddit
  - Three sub-reddit boards: r/NewOrleans, r/Louisiana, and r/GIS

5.2.1.2. Regional Professional Groups

An email announcement was sent to a few regional professional groups, and in each case personal contact was made with leaders. Extended correspondence led to presentations for LaURISA and APA.

- SCAUG (South Central Arc User Group)
- A regional geospatial software user group, based large in the ArcGIS software suite
- LaURISA
  - Louisiana chapter of a North American professional GIS user group
- APA – Metro New Orleans Section
  - Regional chapter of the American Planning Association

5.2.1.3. Historical Societies and Other Heritage Organizations Around the State

An email announcement was sent to each of the following organizations, located across the state of Louisiana: The Cammie G. Research Center (NSU), Cane River National Heritage Area, City of New Orleans GIS Dept., Historic New Orleans Collection, Louisiana Historical Society, Louisiana Trust for Historic Preservation, Midlo Center (UNO), New Orleans Public Library Archives and Special Collections, Preservation Resource Center (New Orleans), Shreveport HPC, Tulane Special Collections, and West Baton Rouge Historical Association.

Generally, contact information was pulled from websites for these groups and the generic stock email was sent.

5.2.1.4. Individuals Researchers and Hobbyists

A number of individuals who had expressed interest in the project, or are known to work locally in relevant fields, were contacted via email. For example, some professors at local universities—LSU of course, but also University of New Orleans and Tulane—and a few Twitter users who are local history enthusiasts in the New Orleans region. Further, this author has a many professional contacts and where relevant interests overlapped these individuals were added to the initial email announcement.
5.2.2. Announcements and Engagement

Not all public announcements about the project took place on the same day. This was not entirely intentional, but generally seemed appropriate out of caution for the untested nature of the platform. However, these staggered announcements turned out to be a boon for analysis, because, generally speaking, each one can be linked to activity of all types on the site—whether visitorship (as measured by Google Analytics), new registrations, or georeferencing activity. Figure 5.2 presents the official user registration numbers on a timeline, and Table 5.1 lists the most prominent outreach efforts, linking them to approximate numbers for public engagement with the site as measured by visitorship and new registrations. A quick comparison between these two datasets clearly shows a correlation, and it will be helpful to have this information in mind while considering the results presented in later sections.
# Table 5.1. Chronological list of prominent outreach efforts.

<table>
<thead>
<tr>
<th>Date</th>
<th>Type</th>
<th>Target</th>
<th>Vis.*</th>
<th>Reg.**</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 2</td>
<td>Presentation</td>
<td>Geo4Lib Community Day – Lightning Talk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb 3</td>
<td>Email</td>
<td>Initial announcement to organizations, historical societies, and individuals.</td>
<td>~75</td>
<td>~10</td>
<td></td>
</tr>
<tr>
<td>Feb 8</td>
<td>Post</td>
<td>Abandoned Louisiana</td>
<td>~170</td>
<td>~30</td>
<td></td>
</tr>
<tr>
<td>Feb 13</td>
<td>Email</td>
<td>LSU SLIS listserv</td>
<td>~55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb 17</td>
<td>Email</td>
<td>SCAUG mailing list</td>
<td>~250</td>
<td>~5</td>
<td></td>
</tr>
<tr>
<td>Mar 24</td>
<td>Post</td>
<td>Nola Devs Slack group</td>
<td>~20</td>
<td>~2</td>
<td></td>
</tr>
<tr>
<td>Mar 24</td>
<td>Post</td>
<td>The Spatial Community Slack groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 29</td>
<td>Post</td>
<td>r/NewOrleans sub-Reddit</td>
<td>~90</td>
<td>~6</td>
<td></td>
</tr>
<tr>
<td>Mar 29</td>
<td>Email</td>
<td>Announcement of two-month extension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 30</td>
<td>Presentation</td>
<td>Open information session</td>
<td></td>
<td></td>
<td>3 attendees</td>
</tr>
<tr>
<td>Mar 31</td>
<td>Presentation</td>
<td>Open information session</td>
<td></td>
<td></td>
<td>3 attendees</td>
</tr>
<tr>
<td>Apr 12</td>
<td>Post</td>
<td>r/Louisiana sub-Reddit</td>
<td>~2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr 13</td>
<td>Presentation</td>
<td>LaURISA lunch and learn</td>
<td>~3</td>
<td>~20 attendees</td>
<td></td>
</tr>
<tr>
<td>Apr 14</td>
<td>Post</td>
<td>r/GIS sub-Reddit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 11</td>
<td>Presentation</td>
<td>HNOC Staff</td>
<td>~2</td>
<td>18 attendees</td>
<td></td>
</tr>
<tr>
<td>May 27</td>
<td>Presentation</td>
<td>APA -Metro New Orleans Section</td>
<td></td>
<td>28 attendees</td>
<td></td>
</tr>
<tr>
<td>May 31</td>
<td>Post</td>
<td>Abandoned Louisiana</td>
<td>45</td>
<td>1</td>
<td>8 shares 2 comments</td>
</tr>
</tbody>
</table>

*Number of site visitors on this day—perhaps not entirely attributable to the announcement, but relevant. **An estimate of registrants as result from this day’s announcement (i.e. registered day of or soon thereafter).
5.3. Georeferencing Activity

5.3.1. Sessions

The work session paradigm outlined in section 4.2.2.2 enables very granular analysis of how users performed tasks on the platform. For example, each time someone prepared a document, a new PrepSession object was created to record the action and its accompanying input data—if five PrepSessions were created on a day, then five documents had been prepared. Additionally, the extra, Sanborn maps-specific step of “loading the volume” can be shown alongside these sessions to give a complete picture of all georeferencing activity performed. This allows us to visualize the all work completed on the platform, as shown in Figure 5.3. One interesting thing that the figure clearly shows is that on certain days many sessions were created for a single type of georeferencing task. This seems especially common for preparation sessions. A few chunks of trim sessions are also apparent, and these times correspond to a lot of work being performed by a small number of users on New Orleans, the trimming necessary to create a clean mosaic of the French Quarter.

Figure 5.3. Full timeline of all georeferencing activity. Note: Two volumes were loaded prior to the public participation period in order to create the tutorial, hence 265 total loads, not 267.
5.3.2. Session Duration

Another benefit of using the session model is that it allows us to calculate how long a user spent using the interface for each step. This calculation is made by subtracting the “created” timestamp (which records when the session object was created, i.e. when the user entered the tool interface) from the “run” timestamp (which records the time when the user submitted their work), and it was a fortunate side-effect of a significant code refactor that was deployed to the system on March 14. This refactor consolidated all session models into the `SessionBase` inheritance pattern, standardizing the timestamp data recorded for each one. Accordingly, any sessions performed before March 14 are not reflected in the following analysis. However, even with this limitation enough sessions are included to reveal clear usage patterns to help us further understand how users interacted with the site.

Figure 5.4 charts the length of all preparation sessions, and a clear distinction can be seen between the length of sessions for documents that required a split operation and those that did not. All together, only a few seconds were needed for non-split sessions, with split sessions

![Chart showing session length](image)

*Figure 5.4. Duration of preparation sessions.*

\[n=834, \text{mean}=1:07, \text{median}=0:36, \text{mode}=0:03, \text{total time}=15:44:02\]
generally taking under one minute.\textsuperscript{46} In all, users spent a combined 15 hours and 44 minutes performing preparation sessions.

Similarly, charting georeferencing sessions by duration shows the task to have generally taken under two minutes (Figure 5.5), though further analysis could be expected to reveal that the more control points a user added, the longer the process took. Combined, users spent a little over 50 hours on this task.

![Georeference Session Length](image)

Figure 5.5. Duration of georeferencing sessions.

Applying the same treatment to trim sessions, we first notice the relatively low amount of sessions as compared to other tasks. This is to be expected as trimming was always presented as an optional step. Beyond that, however, a new method for holistically trimming all layers for a volume at once was conceived of during the second half of the project. With plans to implement that method in mind after the public participation period, the importance of trimming layers was downplayed in later presentations and instruction about the site. The majority of trim sessions lasted between one and two minutes, and users spent a total of four hours on this task (Figure 5.6).

\textsuperscript{46} One possible explanation for the spikes in frequency of non-split session durations around the 1:00 and 1:30 mark is that users may open multiple browser tabs at once (starting many sessions simultaneously) and then make their way through each tab. This would cause the later tabs to be open for far longer than it took to complete the step within them.
5.3.2. Volume-level Activity

Considering Sanborn volumes as a whole, all but thirteen of the original 267 had been at least partially georeferenced, while 46 were fully georeferenced. Remember, some of these volumes had only one sheet, so “fully” georeferencing them would be exceptionally quick. On the other hand, it is quite common that a piece of a sheet will depict a standalone building or cluster of buildings like a sawmill that is located too far away from present-day buildings—perhaps even far outside of the key map—to be georeferenced, so “fully” georeferencing these volumes may be impossible. Similarly, each of the four New Orleans volumes have a text index that takes up an entire sheet, so that sheet will never be prepared, let alone georeferenced.

One especially interesting aspect of the work session paradigm is that we can track the completion of all work performed on a given volume over time. To illustrate this, Figure 5.7 charts all of the sessions that were performed on Documents and Layers from New Orleans, 1885, vol. 2. The initial load date of this volume was February 4, the day after the public announcement of the project. However, no activity was undertaken on it for almost a month, at which point every sheet was prepared all on the same day. A week or two after this mass preparation, georeferencing began and trickled in over the course of a few weeks. Two
significant days of trimming also took place, largely coinciding with outreach efforts as mentioned above.

5.3.3. User Activity

Of the 66 users who registered on the site, 36 performed at least one georeferencing task or volume load—i.e. they contributed in some way to the results described thus far. Table 5.2 lists all 36 of these users, ordered by the date they joined the site (cross-referencing this list with Table 5.1 gives a decent indication of how these users learned about the site). Even a cursory glance at the table reveals a very significant disparity in how and how much different people participated. The top five users—WallyKitty, CaroOkay, sb730_NOLA, eliotj, and nolanc—accounted for 96 percent of all sessions, and the top one user, WallyKitty, accounted for 80 percent of all sessions. Considering the distribution of different types of sessions across users, a relatively large proportion of participants, 14 or a little under half, loaded a volume (or two) but never completed any other actions like preparing or georeferencing sheets. On the other hand, a fair number of other users never loaded a volume, but prepared and georeferenced sheets from those that others had already loaded. The differences between these groups of users will be discussed in the following chapter.
Table 5.2. Users that performed one or more task, ordered by date joined. Compare with Table 5.1 to infer how each user learned about the project. User names have been abbreviated here where they appear to contain full first and last names.

<table>
<thead>
<tr>
<th>User</th>
<th>Joined</th>
<th>Load</th>
<th>Prep</th>
<th>Georef</th>
<th>Trim</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>asluyter</td>
<td>2022-02-03</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>BeckyS</td>
<td>2022-02-03</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>eliotj</td>
<td>2022-02-03</td>
<td>13</td>
<td>29</td>
<td>64</td>
<td>57</td>
<td>163</td>
</tr>
<tr>
<td>nellysun</td>
<td>2022-02-03</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>WallyKitty</td>
<td>2022-02-03</td>
<td>157</td>
<td>1462</td>
<td>1190</td>
<td>28</td>
<td>2837</td>
</tr>
<tr>
<td>brendan</td>
<td>2022-02-04</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>jmcenew</td>
<td>2022-02-04</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>jston</td>
<td>2022-02-04</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>lewinkler</td>
<td>2022-02-05</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>dspreyer</td>
<td>2022-02-08</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>gardnerqd</td>
<td>2022-02-08</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>gigi1960</td>
<td>2022-02-08</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>hurricanesara</td>
<td>2022-02-08</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>jstear</td>
<td>2022-02-08</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>karl</td>
<td>2022-02-08</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Lizartin</td>
<td>2022-02-08</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>trent</td>
<td>2022-02-08</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>EBreaux</td>
<td>2022-02-09</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>lakeeshaj98</td>
<td>2022-09-09</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ddlafleur</td>
<td>2022-02-10</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>jklay</td>
<td>2022-02-10</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>CC</td>
<td>2022-02-12</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>EVW</td>
<td>2022-02-15</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>EmilyF</td>
<td>2022-02-16</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ehoward</td>
<td>2022-02-17</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>nqe1</td>
<td>2022-02-19</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Fullthrottlesuspension</td>
<td>2022-02-27</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>estancioff</td>
<td>2022-03-19</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>nolanc</td>
<td>2022-03-23</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>14</td>
<td>41</td>
</tr>
<tr>
<td>sb730_NOLA</td>
<td>2022-03-25</td>
<td>7</td>
<td>18</td>
<td>80</td>
<td>81</td>
<td>186</td>
</tr>
<tr>
<td>skunkman56</td>
<td>2022-03-29</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>barryg</td>
<td>2022-03-30</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Nellieb</td>
<td>2022-03-30</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>CaroOkay</td>
<td>2022-03-31</td>
<td>44</td>
<td>51</td>
<td>101</td>
<td>1</td>
<td>197</td>
</tr>
<tr>
<td>jwgillie</td>
<td>2022-04-13</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>kevinjm</td>
<td>2022-04-13</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

5.4. User Survey Questionnaire

Eleven respondents filled out the user survey questionnaire, and of these nine had registered and participated to some degree in the project. Basic information about the
respondents showed eight to be Louisiana residents, and, accordingly, in answering whether they viewed maps on the website that were of personal interest (“your home, where you work…”) these eight all said “yes.” On the other hand, not all of them georeferenced maps of personal interest, indicating that some may have been less technologically inclined while still having a closer personal interest in the materials.

Participants were asked how they heard about the project, and many responded saying they had been contacted directly as part of the email blast. A few others heard via Facebook or Reddit. One reported he had learned of the project because his partner encountered a half-page flyers that had been placed at a co-working space in New Orleans. When answering what social media or professional groups they were part of, a surprising amount mentioned Reddit—surprising only because these users did not first hear about the project on Reddit, but would have seen announcements on that site nonetheless.

When asked about their interest in participating, the survey presented users with eight different options, derived from Bonacchi et al. (2019, 176), and were allowed to select as many as desired. Ten users checked the box for “General interest and curiosity,” and 8 checked the boxes for “Enjoyment” and “Contribute to knowledge production” (Table 5.3). That only one user selected “Identity and self-definition” may seem at odds with the fact that so many indicated they looked at maps of personal relevance to them, but it may simply be two different ways of expressing the same thing.
Table 5.3. Summary of user-reported interests

<table>
<thead>
<tr>
<th>Reason for Interest</th>
<th>User Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>General interest and curiosity</td>
<td>10</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>8</td>
</tr>
<tr>
<td>Contribute to knowledge production</td>
<td>8</td>
</tr>
<tr>
<td>Learning about history through maps</td>
<td>7</td>
</tr>
<tr>
<td>Helping out</td>
<td>7</td>
</tr>
<tr>
<td>Giving back to/connecting with the Sanborn map collection</td>
<td>6</td>
</tr>
<tr>
<td>Skill building or career development</td>
<td>3</td>
</tr>
<tr>
<td>Identity and self-definition</td>
<td>1</td>
</tr>
</tbody>
</table>

Overall, the low response was somewhat disappointing, but perhaps unsurprising; in retrospect, extending the public participation period from two months to four likely reduced the number of people with interest in filling out a survey at the very end. Perhaps a better way to distribute the survey would have been to make it available from the outset, making it easy for people to fill it out after their first encounter the site. This approach may have a drawback of not reflecting a user’s entire experience with the platform, but that may be a reasonable trade-off for higher participation. Even though the survey was only ever considered a secondary data collection method after the raw user-generated usage statistics presented above, they do offer some further qualitative insights that the other statistics do not.
Chapter 6. Discussion

There are many aspects of this project that warrant discussion and this chapter will present them in two sections. The first is a discussion of the software platform itself—what worked well, what did not, and what could be improved in the future. The second section will discuss the trends that emerged around user engagement, and propose some ideas for future research on this topic. The next and final chapter concludes this thesis, pulling these discussions together and putting them in broader context.

6.1. Considering the Platform

6.1.1. Pursuing the Archival Commons

The first thing to evaluate is the degree to which this crowdsourced georeferencing platform did, in fact, reflect the ideas of an archival commons. A few clear positives stand out. For example, the utilization of multiple, discrete georeferencing steps allowed for a high level of asynchronous collaboration between users, where one could begin working on a volume, and others could finish it days or weeks later. Additionally, recording work sessions for each user action creates a lineage that can be viewed and summarized. Similarly, attaching usernames to each GCP allows for an even more granular user narrative to be constructed, when a document was georeferenced multiple times by different users (one could imagine an administrative or staff user performing corrective updates on existing crowdsourced GCPs). The foundation for these features sits within the underlying principles used for the georeference extension, and also correspond to aspects of Anderson and Allen’s original vision.
Certain aspects of GeoNode were considered a good match for an archival commons, but participants did not utilize them as much as they could have. For example, the built-in user messaging system could allow various users to work together to coordinate their efforts (one may load a volume, and ask another to begin georeferencing it, for example), but it was never utilized in this manner. Also, GeoNode allows users to rate, favorite, and share individual Documents, Layers, and Web Maps. While especially attractive in theory, this functionality became less useful as the “Volume Summary” page of LOC Insurance Map app became the more important access point for exploration, rather than the default item detail pages for individual Documents or Layers. Further, the volume summary came to aggregate all layers for a volume, a function that was initially imagined for GeoNode’s Web Map resources. This page, as a non-native GeoNode part of the application, fell outside of these built-in utilities and no time and resources were available to integrate it. On the other hand, on a few occasions users did take advantage of the ability to comment on a Document resource, typically saying something like “There are no identifying markings to make this match anywhere.” Recognizing this, it may be the case that adding notes to each GCP (as was the original design) is an unintuitive and unnecessary capability, as long as users can comment on the historical document as a whole.

As for the visual implementation of all these features throughout the site, there are only a few examples to be found, ultimately leaving a lot to be desired. By default, GeoNode allows the assignment of an “Owner” to each resource. This ownership was used to track a user’s interactions with resources—if they georeferenced a Document, they would be set as the Owner of the resulting Layer. Correspondingly, on a user’s profile, a list of all their Documents and Layers could be found. There is also a default GeoNode “Recent Activity” page, which shows
the creation of new Documents and Layers. These pieces of the application were obviously not
designed with the georeferencing extension in mind, but they did provide some aspect of the
“provenance of the narrative” tenet of the archival commons (Anderson and Allen 2009, 395).

The Georeference info panel in the Document and Layer detail pages was designed more
specifically to report the lineage of a resource, and it does begin to create a user narrative. A
user’s name and profile are attached to all sessions that have been performed on a resource, so
from this panel it is easy to see who has contributed to it. However, this should be seen as only a
start. Consider the charts and graphs in the previous chapter: these show the type of user-
centered data that are available. A user’s profile could present them with a list of all the tasks
they had completed, or the front page of the website could use this data to show an interactive
chart of recent user activity. (Map Warper has something like this, and it is a very nice feature.)
Again, time was a limiting factor in pursuing these types of enhancements, but another deterrent
is the fact that updating existing pages in GeoNode requires an overwrite of an existing template
and while possible to do, it is best to keep this kind of alteration as rare as possible. One idea in
the future could add these components to the default GeoNode core templates, and only show
them if the georeferencing extension has been installed and is in use.

6.1.2. Gamifying the System

Gamification is commonly used in crowdsourcing of all kinds, and some aspects of the
archival commons—like the reputation of the agents and user-driven narratives—feed directly
into the basis of gamification. Likewise, this platform is primed to be enhanced to make
participation on it more game-like, given its foundation in a model that supports much of the data
needed to support. For example, a common strategy is to track and publish user statistics like a
“leaderboard,” or perhaps have badges that can be attached to a user once they have achieved certain milestones. A glance at the figures throughout the previous chapter show that there is more than enough information available to create these types of visualizations. In fact, there were already two features of the site that were slightly game-like. For one, on the homepage a counter displayed the number of volumes that had been started out of the total set in the system (e.g. “260/267 Volumes Started”). Also, very near the end of the public participation period, a new column was added to the table that displays all started volumes which uses a small pie chart to illustrate the georeferencing completion for all documents in each volume. This was as much for the author’s need to summarize the work as it was to encourage others to participate, but its addition certainly gives some idea of what is possible.

Ultimately, gamification was not strongly pursued at the outset due to the potential for creating misplaced incentives. What exactly should be incentivized, anyway? Number of Documents prepared? Number of Documents georeferenced? Number of Layers trimmed? Number of other user’s documents georeferenced? Elapsed time spent georeferencing? There is much room for further research here, so it is perhaps best to consider the outcome of this georeferencing effort as minimum of sorts—it was achieved with relatively little attention paid to using conventional incentive structures.

6.1.3. Weighing GeoNode

GeoNode is a heavy-duty application, with capabilities far beyond those specifically needed to support this particular project. Installing, configuring, running, and maintaining it proved to be challenging for this author—stretching his technical abilities considerably! Paying out-of-pocket to run the server was doable in the short-term, but the expense was higher than
anticipated, especially due to server capacity upgrades that were found to be necessary after the beginning of the project. A lot of trial and error went into the georeference extension’s integration into GeoNode, and ultimately, after working through a number of issues, the final strategy is sound but for the specific method that “fakes” an upload of the georeferenced GeoTIFF in order to get it registered in GeoNode as a Layer (and served through GeoServer). This process caused many problems, due largely to signals that triggered extra background processes which significantly slowed GeoServer down, and in some cases crashed it. A likely better approach would be to interface with the GeoServer REST directly to add the GeoTIFF as a layer, and then register that layer in GeoNode.

The integration between GeoNode and GeoServer is complex, largely due to the need to support unified user account authentication and authorizations between the two apps. Data access permissions were a non-issue in the design of this project, due to all of the maps being in the public domain, and no extra “staff” or “admin” accounts were needed save for the single superuser administrator controlled by this author. Thus, having to deal with the complexities of user permissions within the GeoNode/GeoServer architecture often felt unnecessarily burdensome.

Ultimately, the most difficult to handle aspect of the set up, which is certainly due in part to this author’s relative lack of experience, was that upon resource creation or modification certain background processes were triggered within GeoNode, and if too many of these overlapped they would (somehow) crash GeoServer. When this happened, no georeferenced layers would show up on web maps (e.g. the volume summary Map Preview would not show any overlays). Restarting GeoServer proved to be the only remedy in such situations. While these
issues were eventually tracked down and handled (for the most part), they cut significantly into the available time to make other, project-specific enhancements to the platform. Relatedly, as the number of layers grew into the several hundreds, some of these background processes would take longer and longer due to the need to iterate the GeoServer’s layer catalog. No good solution was found for this, and it would have to be reexamined.

GeoNode, though solid in its own right, is probably not the best solution for this platform were it to be pushed further as a purpose-built, Sanborn map georeferencing application. That said, GeoNode continues to be used by other institutions as a geodata portal, so having a georeferencing extension available for it is just as good an idea as it was at the outset of the project. In fact, a new cultural heritage-focused installation of GeoNode is currently in development, to serve as the basis for a digital historical atlas of Poland (Słomska-Przech et al. 2021), and installations such as this would seem to be especially well-suited for the inclusion of georeferencing capabilities. New features of GeoNode also make it more attractive for cultural heritage applications, for example a new component called “GeoStories” allows users to author single-page, scrolling narratives that incorporate layers and maps from within the system. Of course, this capability would have been especially intriguing for people exploring historical Sanborn maps, but trouble with the component (and its initial addition to GeoNode only occurred during this development process) left that avenue unexplored.

6.1.4. Handling Sanborn Maps

As should be apparent by now, but will be reiterated, the Sanborn map collection was a wonderful fit for a project like this, and is significant enough to warrant much more attention. In some ways, catering to this collection ended up taking precedent over the creation of a generic
GeoNode extension, because its internal structure and magnificently detailed content provided so much to work with. Additionally, this project only scratched the surface of the collection as a whole, so there are plenty more maps for future iterations of this work to engage with. With that in mind, a few reflections here will express some lessons learned and ideas for future improvement in this process.

First, there is immense strength in handling Sanborn maps as cohesive volumes, as opposed to sorting through each individual sheet. The “Volume Summary” page, while conceived of relatively late in the development process, became an exceptionally useful construct because it was not only a place to get a progress report on how many more sheets still needed to be georeferenced, but it also provided the map interface to actually view and interact with all of the content for that volume. Beyond being a good construct for visual organization of the content, it became clear through the course of the project that certain operations may be better handled not on a per-sheet basis, but a per-volume basis. One example was the later addition of the key map directly into the georeferencing interface. This allowed users to quickly locate the sheet they were working in, and was especially helpful in finding small pieces of sheets that were located away from other ones. This aspect could be improved further by allowing users to add other main content layers into the georeferencing interface, perhaps even later editions of the same city’s maps which tend to have more comprehensive coverage. Trimming turned out to be another operation that, while possible to apply on a sheet-by-sheet basis, really makes more sense when performed on all sheets in a volume at once.\footnote{The basis for this idea came from a discussion with a former staff member of the Boston Public Library’s Leventhal Map and Education Center. Their strategy for creating mosaics can be found here: Geotransform Digitized Urban Atlases, \url{https://cartinal.leventhalmap.org/guides/create-urban-atlas-data.html}, last accessed July 7, 2022.} Pursuing this
strategy would allow the boundaries of each trimmed sheet to be exactly contiguous, leaving no gaps between them and greatly easing the process all around.

Beyond the “volume” there are other structural levels to the collection that the platform could further exploit to enhance the exploration of the maps, like the city/town level, the parish/county level, the state level, or the atlas level, a compendium of all volumes that create a continuous coverage of a city in a given time range, like the four New Orleans volumes used in this project published between 1885 and 1893. Each one of these levels should be looked at as an access and discovery point, and provide a landing page that summarizes all of its descendant parts.

It also became clear that new categories were needed to better reflect the status of all various map sheets and partial sheets. For example, the text-only index sheets in New Orleans maps need not remain forever in the “Unprepared” category, they should have their own category (indeed, why not separate the text indexes from every volume into a new category?). Further, and more importantly, a better method for handling the “unfindable” pieces would greatly enhance the way a volume is presented. There are maps of lumber mills or standalone buildings that simply cannot be placed on a modern web map because the landscape has changed too much. Instead of leaving these items in the “Prepared” category, creating a new category for them would be better. On the other hand, it may be easier to locate these pieces if later volumes were georeferenced before the early ones, because the later volumes generally have a much wider geographic coverage, meaning that the location of an isolated building on an 1885 map may sit squarely within a full sheet from 1920.
In all, there are many, many ways to further enhance a project like this to suit the Sanborn collection, its structure, and its content. The ideas set forth at the beginning of this project serve more as a foundation than a final product.

### 6.1.5. Allmaps and IIIF

Relatively early in the development of the project, in November 2020, I began corresponding with Bert Spaan, an independent software developer and researcher who was at the time beginning work on a new project called Allmaps.\(^\text{48}\) While Spaan was (and continues to be) working on web-based georeferencing, his approach is based on the use of the International Image Interoperability Format (IIIF, or “triple-I-F”) image-serving protocol. Nowadays, many online digital collections use a IIIF server to (among other things) power a zoom interface for deep inspection of their graphical materials. However, the same IIIF services can be utilized by third-party applications to pull these images into external viewers. Spaan’s web georeferencing approach is to create a new IIIF specification that holds a set of GCPs and then applies the georeferencing transformation directly in the browser, warping the IIIF tiles and placing them directly onto a web map. This implements georeferencing not as an operation done to an image to create a geospatial dataset, but as an ephemeral transformation of existing IIIF web services, no matter where they come from.

With this in mind, one early idea for this project was to abandon GeoNode as a base and build directly from a clone of Allmaps itself. Of course, GeoNode still had all of the advantages described in earlier chapters, and Allmaps the disadvantages that are described below. However, the most important difference between them at the time was GeoNode/GeoServer’s ability to

---

\(^{48}\) Allmaps, [https://allmaps.org](https://allmaps.org), last accessed July 7, 2022.
serve maps through the WMS protocol, which would be the main method for publishing the data in a way that other people could use it. Allmaps warped the IIIF tiles but relied on client-side code to do so, meaning that georeferencing maps there would not make them available for use elsewhere; publishing a GeoTIFF through GeoServer makes it available everywhere. Interoperability in mind, using Allmaps outright was a non-starter.49

Thus, through the winter and spring of 2020/2021 the planned approach was to insert Spaan’s IIIF-based interface directly into the center of this project’s georeference extension, such that it was fed a IIIF manifest made from a Document in GeoNode. After the georeferencing was complete, the resulting GCPs, having been accumulated into a list based on a specification that Spaan was working on, would be returned to the georeference extension and used to run the warping process, creating the Layer as a GeoTIFF and loading it into GeoNode as planned.

The main benefit that was seen at the time was the reuse of Spaan’s existing code to create the actual georeferencing interface, and potential integration into his Allmaps project in the future. However, these benefits were judged to not outweigh a few significant downsides, especially when considering the amount of work that is always necessary to adapt one code base into another. First, the in-development GCP JSON specification that Spaan’s approach is based on does not provide for the attachment of the username and optional narrative note to individual GCPs—these two properties were seen as necessary to support the extension’s underlying principles. Second, the masking paradigm used by Allmaps at the time was too simplistic to handle the fact that Sanborn map sheets have multiple areas that must be georeferenced individually (this has since been updated), which meant that even if Allmaps were used some sort

---

49 In October of 2021 Spaan released a prototype of the Allmaps TileServer, which, more or less, addresses this need: the tile server does not publish the maps through the WMS protocol, but instead provides an XYZ tile endpoint that can be used in QGIS and most web mapping libraries.
of splitting and trimming interfaces would have to be built anyway. Finally, though the Allmaps model theoretically did provide the possibility for a live preview after a user had added 3 GCPs, it did not have this feature at the time which was seen as necessary for a user-friendly experience.

For this project, then, the work that had been done up to spring of 2021 to integrate with Allmaps was set aside and the interfaces described above were subsequently created from scratch in order to support the specific needs described here. Nevertheless, significant development has taken place on Allmaps in the meantime so there is now great potential for reapproaching integration with it. Advantageously, the design of this project’s georeference extension organizes its data such that it would be possible to repackage and integrate it into whatever specification Allmaps requires.

As should be clear by now, the approach throughout this project is very analog: files are downloaded from the LOC, they are cut into new JPEGs, and they are georeferenced into new GeoTIFFs. In other words, there is a lot of file manipulation involved, and file storage, though cheap, is a serious consideration when talking about large collections. On the other hand, Allmaps takes a much more light-weight, web-based approach. In its initial design, no files were ever downloaded, and no server is even used; image services that are already available through institutional IIIF servers are augmented directly in the browser to produce map overlays. The Allmaps TileServer somewhat changes this model, because the tiles are requested from IIIF images, downloaded, warped on a server according to the provided GCPs, and then served as map tiles.

Whether every aspect of Allmaps would be a beneficial addition or replacement for portions of the architecture here remains to be seen, but, without a doubt, further investigation into its uses is warranted.

---

50 Interestingly, the Allmaps TileServer somewhat changes this model, because the tiles are requested from IIIF images, downloaded, warped on a server according to the provided GCPs, and then served as map tiles.
6.2. Crowd Activity

One goal of the project was to get a wide variety of people involved in the georeferencing process, especially those who would georeference their hometown or areas where they currently live. As the discussion below will show, this goal was met only to a limited degree, but the results clearly point to methods by which it can be achieved more successfully in the future.

6.2.1. Types of User Activity

Judging from the usage statistics and other metrics presented thus far, it seems there are four loose categories of user activity. First, “Single City” users are visitors to the site that do want to see a specific city’s maps, but are not interested in and/or not technically savvy enough to georeference the maps. This category of user activity became apparent after the first post to the Abandoned Louisiana Facebook group, when about 15 people signed up and loaded a single volume, but then went no further with the site. Though direct feedback was not received from these users (none of them filled out a questionnaire), it seems likely that they were quickly deterred by the comparatively technical steps of the georeferencing process. Notably, at the time these users would have first visited the site, so early in the project, there was help content in the documentation, tutorial, and at the top of each georeferencing tool, but there was not yet an instructional video. This is to say that these earliest users of the site would have received less guidance in their first visit than later users did, and, unfortunately, on the whole these users composed the least technically savvy group that was contacted.

Second, a few users fit into a “Capable Passersby” category. These users georeferenced a few Documents, seemingly to try it out, but did not go much further than that. Given this usage pattern, it seems reasonable to consider this group to have very high potential for participating if
they were particularly motivated to do so. Inasmuch as this platform was intended to create a space where people could find and georeference historic Sanborn maps of their choosing, it seems to have served these users just as well as those in the following groups.

Third, the “Enthusiasts” are those users in the top five of overall session activity, who georeferenced dozens of maps. These users may have initially georeferenced maps of a specific area of interest, but also branched out quickly, eventually georeferencing content from 8-14 different volumes. Fourth, we have the “Power Users” who, in this case, really only consisted of the single, top user, WallyKitty. This user’s impact will be described in the next section, their activity accounting for 80 percent of all work sessions performed on the site.

Reflecting on these various types of activity, each one has its own advantages. Perhaps the most important to point out is the first group. While these users may not have contributed significantly to the number of georeferenced documents on the site, their initial load of a single volume actually speaks clearly to their interest in getting a city’s maps to be georeferenced. This user-to-location connection is to be treasured, because it is the most likely to result in organic sharing and further interactions with the content. Also, members of other groups are less likely to have this connection, as they branch out to georeference other content. On the other hand, of course, it is the other groups who will be performing the actual georeferencing, so one good way to join these motivations would be to make the connection between all of these users more apparent in some way.

The great benefit of the Enthusiasts and Power Users, of course, is that they fill in these gaps, creating content that any number of other people may be significantly interested in. On the other hand the Capable Passersby group may be able to help out a beginner, or could even be
tasked with georeferencing a set of maps for a specific research endeavor—their own or someone else’s. Where the Single City group may not have the skills needed to perform the more technical tasks, having the initial “load volume” step actually seems to have been able to give them a place in the ecosystem, as they can initialize a process and “plant a flag,” so to speak, in a city’s Sanborn maps. With regard to the archival commons, this pattern reflects well on the “organic” nature of the system, because it can grow as desired by its users and appear as a “boundless” collection of Sanborn maps.

Of course, while this expansive model exists in theory, some improvements could greatly enhance and expose it for future projects. For example, one idea—which was actually suggested by a user very early on but could not be borne out due to time limitations—is to add a notification system on a per-volume basis. That way, if a user loads a volume but does not georeference any of its content, a notification would be sent to them whenever one of its sheets is georeferenced by someone with more technical prowess. Having such a system in place would surely have been beneficial for bringing people back to this project’s website long after they had, presumably, given up on it.

6.2.2. WallyKitty: The Top 1 User

So what do we make of WallyKitty? This prolific user accounted for 80 percent of all georeferencing tasks performed on the site, and also maintained activity throughout the duration of the public participation period. Figure 6.1 clearly shows the out-sized impact that WallyKitty had on the project. Where the activity of other users more closely reflects the outreach and announcement timeline, WallyKitty worked on georeferencing fairly consistently throughout the course of the project, with an especially large push from mid-April to mid-May. Some of the
massive spikes in February we know to be bulk preparation sessions—indeed by referencing figures in earlier chapters we can even match one spike to the day on which all 57 New Orleans, 1885, vol. 2 sheets were prepared.

WallyKitty did fill out a user survey questionnaire, so we know this user to be a GIS Analyst who does not live in Louisiana. They reported that the work in the project is potentially related to their line of work because “Historic map reviews are a common part of records search efforts for archaeological projects,” and that they did not view or georeference content of a personal interest. Reporting their interest on the survey, the options they included were: learning about history through maps, giving back to/connecting with the Sanborn map collection, general interest and curiosity, enjoyment, helping out, and contributing to knowledge production. In other words, the most active user in the entire project was a foreign powerhouse with sincere enjoyment for the task and a strong desire to participate in the project. Indeed, the great benefit of WallyKitty’s participation was the sheer volume of contribution, which created so much content for other people to explore. Perhaps even more important, WallyKitty performed a disproportionate amount of preparation tasks, often in bulk all on the same day, which allowed other users to go straight to georeferencing. This indicates that making “bulk” preparations—
specifically, allowing the “No Split Needed” option to be quickly applied to many Documents at once—would greatly improve this user’s experience.

Perhaps the best way to consider WallyKitty’s activity is as an example of how a professional, or a GIS company, could use the platform. In all, this user spent about 15 hours preparing sheets and about 32 hours georeferencing them, for a rough total of 47 hours, or about six 8-hour work days. In this time period, about a tenth of Louisiana’s entire Sanborn content was georeferenced. Could five WallyKittys georeference all Louisiana Sanborn maps in about two weeks? It seems entirely possible, specially given that some very significant improvements could speed up bulk processing.

6.2.3. Improving Engagement

One clear finding through the examination of user activity is that a single user can have a disproportionate impact on the site. One way to look at this could be a deliberate search for “power users” in places they are likely to be found (Reddit seems a good place to start, as the majority of survey respondents have accounts on that site, including those that were the heaviest users). However, this project took a more circumspect approach by casting as wide a net as possible, recognizing that a social media post (or stack of flyers) need only be noticed by one interested user to make a big difference.

Another finding was the importance of personal connection. In the initial email blast, two different stock emails were sent to different sets of individuals. This author knew the individual or knew personally of their interest in the project, a short note was prepended to the message. Emails that contained a note were much more likely to receive at least a short response, and sometimes these responses turned in to exchanges that led to further involvement.
Of the social media communities contacted, each had its strengths. Interestingly, many of the survey respondents indicated that they are members on Reddit, even if Reddit was not how they heard about the site. On the other hand, posts to the Abandoned Louisiana Facebook group were shared to at least one other relevant Facebook group, Lake Charles Historic Neighborhoods, which reached a new set of people with a direct connection to a specific community. More posts like this, especially if they could include specific links to the maps for a city of interest (or even better, specific locations within that city), would be very beneficial.

Finally, one good idea brought up by a user early on in the project, but never prioritized and carried out, was to make it much easier for users to submit feedback at any point during the public participation period by adding a simple contact form directly on the website. Of course, the researcher’s email address and other contact information was always available on the site, but only a few participants emailed out-of-the-blue—more often emails were received after mass updates were sent out to the overall contact list.
Chapter 7. Conclusion

As stated at the outset, a vast amount of historical maps is becoming available online through digital institutional collections. Crowdsourcing has often been the answer to how these maps can be georeferenced, a process that is naturally desirable because it transforms them into new materials with new capabilities. While there are some existing web platforms that support this crowd-based participation, this thesis envisions a new approach, attempting to place web-georeferencing within the wider framework of heritage crowdsourcing, especially using the idea of Anderson and Allen’s archival commons to guide certain aspects of the development. The result, described heretofore, was the creation of an extension that fits on top of existing geospatial content management system software (GeoNode), thereby adding capabilities to a mature open source platform that is already in use by many organizations around the world. The conceptual principles of the extension design were 1) georeferencing as metadata creation, not as dataset creation, 2) iterative and collaborative work, and 3) independent stages. This framework created a good foundation that not only allowed users to efficiently perform the work, but also tracked how they engaged with the platform—who contributed which steps to which maps, charting chronological session activity on the site, even how much time it took to georeference each individual document, or work session duration averages on the whole. Though much more could have been done to directly expose this information on the website itself, building the foundation for its collection was the first step.

However, the second objective of this research came to somewhat overshadow the first. In order to conduct a pilot project around this georeferencing extension, extra structure was layered on top of the application to facilitate a connection to the Library of Congress online
Sanborn Map collection and further enhance the experience of exploring and working with those maps. This collection was chosen not only because of its historically significant material, but also because its wide geographic coverage brings with it great potential for connecting with ordinary citizens. To test this potential, the pilot project targeted Sanborn maps across the entire state of Louisiana, 138 different places in all, with the high-minded idea of working toward something like a “georeferencing commons” for the collection.

To summarize the findings from the pilot project, we will refer back to its three main goals.

1) General software evaluation

GeoNode proved to be a challenging system to build from in this particular instance, due to its general size and the complexity of its components. The georeferencing extension did integrate into it well, and with some targeted updates could be added to any existing or future GeoNode implementation. For the specific purposes of hosting a crowdsourced georeferencing project, however, it was unnecessarily cumbersome and resource-intensive.

The design of the georeferencing extension and connection to the LOC took a very traditional approach of downloading and manipulating files on disks, an approach that looks positively old-fashioned when compared to the latest IIIF-based web-georeferencing systems that are in development. On the other hand, there are many other methods for storing and serving map tiles to explore, and, through the course of this project, potential collaborations with Open Historical Map51 and Machines Reading Maps52 have emerged. However, the choice to store georeferencing data and manage user participation in a very granular way makes the actual

georeferencing mechanics (and how they are performed on the server) somewhat secondary to
the system overall and provide ample opportunities for growth and migrations in the future.

With regard to Sanborn maps themselves, the pilot project revealed a number of
important characteristics of the collection, as well as new ways to build from them in the future.
One of the most meaningful realizations was the benefit of approaching georeferencing these
maps in a holistic, per-volume (not per-sheet) manner. Some ways to do this would be
simultaneously trimming adjacent sheets, or including already georeferenced content from the
volume into the georeferencing interface to aid users.

2) Track and understand user engagement

Overall, 66 users registered to participate in the project, and a little over half of them
performed one or more georeferencing tasks. Four groups of users emerged, and perhaps the
most interesting were individuals who signed up after hearing about the project through a
Facebook group, clearly expressed interest in just one or two cities by loading a volume of maps,
but then went no further. Considering the archival commons, these may be the most important
users to get involved, as the “load volume” step allowed them to express a direct interest in (and
probably connection to) specific map material. Not everyone is technically capable of
performing the georeferencing task however—no matter how cleverly constructed the interface
may be. Thus, the possibility of a “sponsorship” model emerges where one user would “load a
volume” and then be notified later when someone else has georeferenced its content.

On the other end of the spectrum, one single user performed 80 percent of all the
georeferencing tasks that were carried out through the pilot project. While user activity
disparities are common in heritage crowdsourcing efforts, it is further interesting to consider this
user perhaps not as a member of the crowd, *per se*, but as representative of the platform’s capabilities when used by a professional. On average, this user georeferenced a map in under two minutes, generating 1,000 layers in the equivalent of about six work days.

One final thing to consider regarding user engagement is that this project was not carried out through a popular institution, nor there was an existing mailing list to use or social media presence… it was one graduate student talking to anyone who would listen, sending emails, and nervously posting on social media. These efforts were significant, but a larger outreach, sponsorship, or backing behind a similar project would undoubtedly yield a greater response.

3) Generate and publish novel new datasets

Of course, we cannot forget the maps themselves. Over the course of the pilot project, users georeferenced 267 different Sanborn maps across Louisiana (either partially or in full), creating over 1,500 layers. The communities in these maps range from small towns like Baker in 1922 (one sheet) to full coverage of New Orleans between the years of 1885 and 1893 (240 sheets). No Sanborn map mosaics exist like these in the state of Louisiana, and, more broadly, it is rare to find digital applications of them across the nation for small and medium-size towns, like this project provided the means to create. At the end of the summer of 2022, the layers will be made publicly available for long-term use through the LSU Atlas geospatial data portal.

In closing, I will reference *Toward a Philosophy of Photography*, a treatise by media critic and philosopher Vilém Flusser (Flusser [1983] 2012). Flusser sees the camera as an apparatus capable of creating only a finite number of distinct photographs, or “technical images,” constrained entirely by the mechanical logic of the camera itself. The camera defines all the possible outcomes of using it, so while talented photographers may stretch the rules of the
apparatus in experimental and beautiful ways, their creations are still bound by its physical nature. With this in mind, to design a new camera is to immediately create every possible photograph that can be produced by it. This paradigm drove my work on this thesis in many ways, especially when working with the Sanborn map collection. The goal was not to georeference a set number of maps, but to create an apparatus around the collection such that the work of georeferencing its content could become akin to Flusser’s idea of people creating photographs: the platform is the camera, the subject is the maps, and the participants are the photographers. Through this approach, a uniformity to this mass georeferencing could emerge that not only facilitates the transformation of a digital collection, but also deepens everyone’s connection to it.
Appendix A. IRB Materials

Determination Letter

TO: Andrew Sluyter
LSUAM | Col of HSS | Geography and Anthropology

FROM: Alex Cohen
Chairman, Institutional Review Board

DATE: 24-Jan-2022

RE: IRBAM-22-0001

TITLE: Creating a Public Space for Georeferencing Historical Fire Insurance Maps

SUBMISSION TYPE: Initial Application

Review Type: Exempt

Risk Factor: Minimal

Review Date: 23-Jan-2022

Status: Approved

Approval Date: 23-Jan-2022

Approval Expiration Date: 22-Jan-2025

Exempt Category: 2b

Requesting Waiver of Informed Consent: Yes

Re-review frequency: Three Years

Number of subjects approved: 250

LSU Proposal Number:

By: Alex Cohen, Chairman

Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*

2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.

3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.

4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
Informed Consent Form

The following informed consent form was presented to users above the registration page on the website, and before the user sign-up form was enabled they were required to 1) reveal it with a click and then 2) scroll down and check a box agreeing to it.

Study Title
"Creating a Public Space for Georeferencing Historical Fire Insurance Maps"

Purpose and Procedures of the Study
Members of the public are invited to engage with historical map archives by participating in a crowdsourcing project. Registered members can access old fire insurance maps of Louisiana communities and use tools on this site to transform the files into interactive web maps. For the duration of the study period, general usage statistics will be collected to better understand user engagement. At the end of the study period a brief, optional, survey will be distributed via to participants in order to learn about their experience and reasons for participating in the project.

Those who do not wish to participate can still view, download, and explore any content on this site.

Participant Inclusion Criteria
Individuals over 18 years of age.

Participant Exclusion Criteria
Individuals under 18 years of age.

Number of Participants
There is no minimum number of expected participants. The maximum number of participants is 250.

Privacy
Results of the study may be published, but no names or identifying information will be included in the publication. Subject identity will remain confidential unless disclosure is required by law.

Risks
This study presents no known risk to participants.

Benefits
Georeferencing old maps transforms historical documents into a digital format that can be used and reused in a wide variety of future research contexts. Participants can expect to gain historical knowledge of the Louisiana communities they are interested in, and all digital content and map data generated through this project will be in the public domain.

Compensation
No financial compensation will be offered to participants.

Voluntary Nature of Study
Participants may choose not to participate or to withdraw from the study at any time without penalty or loss of any benefit to which they might otherwise be entitled.

Location of Study
All participation will take place online at oldinsurancemaps.net, and the user survey will be administered online via Qualtrics.

**Investigator Contacts**
Andrew Sluyter (asluyter@lsu.edu) or Adam Cox (acox42@lsu.edu, 608-606-9928) can be contacted with questions or concerns about this study.

This study has been approved by the LSU Institutional Review Board (IRB). For questions concerning participant rights, please contact the IRB Chair, Alex Cohen, (225) 578-8692, or irb@lsu.edu.

By signing up you are giving consent to participate in this study and to be e-mailed a brief, optional, survey about your participation at a later date.
Appendix B. User Survey Questionnaire

User Questionnaire for LaHMG

The following questionnaire was administered through Qualtrics.

Section One: About You

(all questions optional)

Q1 If you signed up on oldinsurancemaps.net to participate, what was your username?
_________________________________

Q2 Are you a Louisiana resident?
   ○ No (1)
   ○ Yes (2)

Q3 If yes, which parish?
_________________________________

Q4 What is your age range?
   ○ 18-24 (1)
   ○ 25-34 (2)
   ○ 35-44 (3)
   ○ 45-60 (4)
   ○ 61+ (5)

Q5 What is your job/occupation? (you can be as specific or general as you like)
_________________________________

Q6 For which of the following do you often use computers and/or the internet? (check all that apply)
   □ Social media/emails (1)
   □ Online shopping (2)
   □ Spreadsheets/databases (3)
   □ GIS/mapping applications (4)
   □ Web development/programming (5)
Q7 Are you a member of any of the following groups? (check all that apply)

☐ Abandoned Louisiana (Facebook Group) (1)
☐ LaURISA (2)
☐ SCAUG (3)
☐ American Planning Association (4)
☐ Reddit (r/Louisiana, r/NewOrleans, r/gis, etc.) (5)

Section Two: Your Interest in the Project

(all questions optional)

Q8 How did you learn about this project?
_________________________________

Q9 How would you best describe your interest in using oldinsurance maps.net? (check all that apply)

☐ Learning about history through maps (1)
☐ Giving back to/connecting with the Sanborn map collection (2)
☐ General interest and curiosity (3)
☐ Skill building or career development (4)
☐ Enjoyment (5)
☐ Helping out (6)
☐ Contribute to knowledge production (7)
☐ Identity and self-definition (8)
☐ Other: (9) _________________________________

Q10 Are there aspects of this project that are relevant to your line of work?

☐ Yes (1)
☐ No (2)

Q11 If yes, how so?
_________________________________

Q12 In which of the following ways did you interact with the site content?

☐ Viewed maps (1)
☐ Shared maps with others (2)
☐ Created control points to georeference documents (3)
☐ Trimmed layers (4)
☐ Authored new web maps by combining layers (5)
☐ Other: (6) _________________________________

Q13 Did you georeference maps of an area of personal interest?
   *For example: your home, where you work, or the street you grew up on*
   
   ☐ No (1)
   ☐ Yes (2)
   ☐ Wanted to but someone already had! (3)

Q14 Did you explore web maps of an area of personal interest?
   *For example: your home, where you work, or the street you grew up on*
   
   ☐ Yes (1)
   ☐ No (2)

Q15 Do you have any feedback for how to improve the site, or similar efforts in the future?
   ___________________________________________
Appendix C. Session Data Examples

PrepSession.data

{   'split_needed': True,
'cutlines': [   [[2522,0], [2519,8014]],
    [[2521,2176], [6490,2183]]
],
'divisions': [   [[0,0], [0,7650], [2519,7650], [2523,0], [0,0]],
    [[2519,7650], [6450,7650],[6450,2182], [2522,2176], [2521,2176],
     [2519,7650]],
    [[6450,2182], [6450,0], [2523,0], [2521,2176], [2522,2176], [6450,
     2182]]
]
}

This JSON object stores 1) whether a split operation was needed on the Document, 2) each of the cut-lines that were drawn by the user (in this example, two), and 3) the boundaries of the divisions that these cut-lines created (in this example, three). All coordinates are based on the dimensions of the original image.

GeorefSession.data

{   'epsg': 3857,
'transformation': 'poly1',
'gcps': {   'type': 'FeatureCollection',
'features': [   {   'type': 'Feature',
'geometry': {   'type': 'Point',
'coordinates': [-92.4539626855611, 31.312901950623868]
},
'properties': {   'id': 'f2210d84-c6f1-43d7-9800-daa2fee64971',
'note': '',
'image': [2428, 1610],
'listId': 1,
'username': 'acfc'
}]
}
This JSON object consists of 1) the EPSG code that was in use when the GCPs were recorded, 2) the GDAL algorithm that was used to warp the image with these points, and 3) a GeoJSON representation of all GCPs. Note that the GCP geometry coordinates are stored as longitude, latitude (per GeoJSON spec), and within the properties for each point an “image” attribute stores the corresponding pixel coordinates for this GCP. Other properties include the name of the user, an optional note, and internal identifiers.

**TrimSession.data**

```json
{
  'mask_ewkt': 'SRID=3857;POLYGON ((-10026799.2 3494568.3, -10027094.3 3494899.7, -10027290.3 3494745.9, -10027016.7 3494389.9, -10026799.2 3494568.3))'
}
```

The data stored for a trim session is an EWKT (Extended Well-Known Text) representation of the layer mask polygon.

**Note**

An astute geospatial professional will perhaps notice, with dismay, that each of the structures above uses a different geospatial format for representing geometries—a simple list of coordinates, GeoJSON, and EWKT. GeoJSON is a clear choice for the set of GCPs in a GeorefSession, because this fits the specification in development by the IIIF community. However, proper GeoJSON requires that coordinates be represented with the World Geodetic...
System 1984 (WGS 84) (EPSG 4326) using longitude, latitude. This means that the document-based coordinates representing cut-lines and divisions cannot (or should not) be stored in a GeoJSON format. Similarly, the mask polygon is always drawn and displayed in the WGS 84 Web Mercator (EPSG 3857) coordinate system, so it seemed reasonable to store it in that coordinate system, and use a format that can be easily read and written.

That said, this appendix shows only what is, not necessarily what should be.

---

References


121


Vita

Adam Cox, a native of Viroqua, Wisconsin, received his Bachelor’s degree in Anthropology from Lawrence University in 2011. Since that time, he has worked for the National Park Service and as an independent consultant in the geospatial industry, often on map-based web applications for heritage institutions. He is a strong advocate for open source software, especially its application within the wider geographic information systems and information science communities. He entered the LSU Geography & Anthropology department in 2019, and the LSU School of Library & Information Science the following year. He expects to complete his dual degree program—Master of Science in Geography and Master of Library & Information Science—in August, 2022.