Graduate Theses, Dissertations, and Problem Reports

2020

The Spatial Historian: Creating a Spatially Aware Historical Research System

JAMES P. SCHINDLING
jpschindling@mail.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Part of the Digital Humanities Commons, and the Other History Commons

Graduate Theses, Dissertations, and Problem Reports

2020

The Spatial Historian: Creating a Spatially Aware Historical Research System

JAMES P. SCHINDLING

The Spatial Historian:
Creating a Spatially Aware Historical Research System

James P. Schindling

Dissertation submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Geography

Trevor Harris, Ph.D. (Chair)
Melissa Bingmann, Ph.D.
Jamison Conley, Ph.D.
Kenneth Martis, Ph.D.
Matthew Vester, Ph.D.

Department of Geology and Geography

Morgantown, West Virginia 2020

ABSTRACT
The Spatial Historian:
Creating a Spatially Aware Historical Research System

James P. Schindling

The intent of this study is to design a geospatial information system capable of facilitating the extraction and analysis of the fragmentary snapshots of history contained in hand-written historical documents. This customized system necessarily bypasses off-the-shelf GIS in order to support these unstructured primary historical research materials and bring long dormant spatial stories previously hidden in archives, libraries, and other documentary storage locations to life. The software platform discussed here integrates the tasks of information extraction, data management, and analysis while simultaneously giving primary emphasis to supporting the spatial and humanistic analysis and interpretation of the data contents. The premise of this research study is that by integrating the collection of data, the extraction of content, and the analysis of information from what has traditionally been post-data collection analysis and research process, more efficient processing and more effective historical research can be achieved.

DEDICATION

This thesis is dedicated to Lugene, Kyle, and Casey.

…and to all those who are not too old to learn.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - Introduction

On January 20th, 1590, the baptism of Mateo was recorded in a six-line hand-written entry in a register in the *Sagrada Catedral de San Christobal de La Habana* in Havana, Cuba. No surname was given for Mateo nor his parents Mateo and Ysabel – only the descriptive *Criollo*, and that they were *Negroes* owned by Alonso Çuares of Toledo. Presiding over the ceremony was Juan Ponce Roman. Also present were Fancisco Surdebiera, the younger Mateo's godfather, and Leonor de Avellaneda, the godmother and slave of Baltasar de Biera. The signature of Juan Ponce Roman dominated the entry, consuming as much space as three lines of the record itself. With the ink dried and the page turned for the next in the series of entries, it is possible that the record of the baptism of Mateo, the Negro slave, had lain unseen for over four centuries. This is one among thousands of examples of the small slices of history residing on shelves and in basements of houses of worship and local archives – sometimes known, sometimes unknown – waiting to be opened and placed within a larger narrative of the past.

The intent of this study is to design a geospatial system capable of facilitating the extraction and analysis of these fragmentary snapshots of history contained in hand-written historical documents. For several reasons, as elaborated below, this customized system necessarily bypasses off-the-shelf GIS in order to support these unstructured primary historical research materials and bring long dormant spatial stories previously hidden in archives, libraries, and other documentary storage locations to life. This software platform integrates the tasks of information extraction, data management, and analysis while simultaneously giving primary emphasis to supporting the spatial and humanistic analysis and interpretation of the data contents.

The importance of space in historical studies is reflected in the advent of the Spatial Turn – often referred to with the qualifying prefix "so called". The Spatial Turn recognizes that the examination of history from a spatial-temporal perspective offers the potential of new and innovative lines of historical inquiry and thereby enable new insights into the historical past. Two manifestations of the Spatial Turn can be discerned from the literature. The first relates to discussions of the role of space as lived places with historical significance. A second manifestation of the Spatial Turn is as a method for understanding the control, management, and manipulation of society (Foucault, 1984; Harvey D. W., 1984; Lefebvre, 1991; Soja, 1989; Tuan, 1979). Motivated by perceived political, economic, religious and social injustices, humanist scholars write in often noble terms describing the ways in which spaces are manipulated by institutional forces with the goal of suppressing, subjugating, and otherwise controlling populations for their own advancement. This second manifestation especially relates to criticism made about the role of GIS and the application of computer technology captured by the term digital humanities.

While the Spatial Turn has been heralded as a sea change in social theory, it has arguably produced little recognizable change in mainstream historical writing. With the exception of micro-historical research, where the intimacy of a locality is integral to the subject, the focus among historians has largely remained on detailing a narrative of events and personal interactions. On the other hand, the development of historical GIS (HGIS), GIScience, and spatial analysis as adjunct historical research methodologies have been applied to a broad range of subject areas (Harris T. M., 2002; Knowles A. K., 2008; Owens, 2007). When used in conjunction with advances in spatial data repositories and spatial data availability, including remotely sensed satellite imagery and LiDAR data, such systems promise considerable advances in archaeological and historical analysis (Garrison, Houston, & Firpi, 2019; Moyes & Montgomery, 2019; Schindling & Gibbes, 2014).

Thus, this paper discusses the conception and development of the *Spatial Historian*, a customized geospatial platform developed by this author that supports extensive historical data management services including text transcription, content normalization and translation; the ability to record, track, and geotag people and events to places through coordinate systems, placenames, or toponyms; integrated spatial data exploration and spatial analysis; support for collaborative research with remote system access and multiple research projects with cross-project integration; and finally the generation of industry standard metadata to support export to digital archives and data curation. This study and resulting software system seek to incorporate geospatial capabilities into the historical conversation that extends from data collection to study findings

The Spatial Humanities

Technological developments as outlined above are based on understanding gained from the conceptual underpinning of the knowledge domains that they are designed to support. Geographical technologies must also incorporate data, findings, and insight gained from existing scholarly research and prior studies. As such, a knowledge-base of three broad subject areas are important in this study to understand the relationship between the goals of historical researchers and the ways in which geospatial technology, as developed here, can be most effectively applied. These three subject areas include the spatial humanities and deep mapping, historical GIS, and the curation of digital data, and in particular, geo-spatial data.

The spatial humanities promote the role of space in the analysis and interpretation of humanistic studies and has been heavily influenced through the application of geographical technologies (Bodenhamer, Corrigan, & Harris, 2010; Gregory, Donaldson, Murrieta-Flores, & Rayson, 2015). The spatial humanities seek to understand how space is an active agent in the lives

of people and how people in turn affect their surroundings. In the examination of historical records, place and movement through space are essential forces in historical narrative, for space is not only the stage on which events take place, it also plays a leading role in human behavior. Integral to the spatial humanities are the ways in which space has been examined and interpreted within the humanities. Spatial history brings the concepts of the spatial humanities to bear on historical research by exploring the ways in which historians incorporate an understanding of place within their research and how the concept of place in historical studies has evolved (White, 2010). The idea of 'place', in the sense of both a description of location and an internalized understanding of one's surroundings, varies over time and across cultures. These variations can be detected in linguistic forms, physical artifacts, and through the extraction of spatial content embedded in historical materials.

A long-standing approach to gaining a geographical context of past events has been to supplement text narrative with maps. An excellent example of this is the recent publication of an annotated version of Herodotus' Histories that includes 137 maps that show the location of every significant place or event described in the original text (Strassler, 2009). However, gaining a greater sense of historical events entails not only building gazetteers and maps but developing an understanding of the contemporary geographic and environmental conditions that may have influenced the actors. An example of this can be seen in *The Mediterranean and the Mediterranean world in the age of Phillip II* (Braudel, 1995), where Braudel explores a 'total history' approach to the ways in which the land, sea, and climate affected cultural development and social interaction across the region.

More recently, the development of HGIS as an outgrowth of traditional GIS and historical geography has been explored to identify new ways of looking at spatio-historical information and

to draw out patterns of actions and events. HGIS is playing an ever increasing role in this historical mapping process (Harris T. M., 2002; Knowles A. K., 2008). For Knowles, for example, General Lee's seemingly uncharacteristic order to have General Pickett charge directly into the strength of the Union army at the Civil War battle of Gettysburg caused her to wonder if there may have been other circumstances that had gone unnoticed. Using GIS, Knowles built a virtual reconstruction of the battlefield at the time and used viewshed analysis to determine if Lee's view of the field from the cupola of the Lutheran Seminary may have led him to misread the strength of the Union army there. Her research showed that there is a case to be made that the terrain and vegetation at the time may have obscured his view and therefore influenced his decision. Harris used similar GIS analysis to investigate the seemingly unplanned placement of habitations and farmsteads in relation to the location of the Danebury hillfort in England. To accomplish this, he developed a friction surface as an indicator of the time, effort, and resources required to traverse the surrounding landscape. The result of the analysis showed that, even though the spatial distances from the farmsteads to the safety of the hillfort may have varied, the time that it took to travel those distances was within a relatively fixed range. These examples of spatial analysis and the application of HGIS provide new insights into how the spatial context is an integral component of human activity. The positive impact of GIS technology can also be seen by comparing the sophisticated analytical maps produced by Harris and Knowles to the hand-drawn black-and-white maps of the earlier work by Braudel.

HGIS, however, is still in its infancy with its existence predicated on accessibility to computer-based mapping capability that has only reached a degree of maturity within the past decade. Even now GIS capability continues to develop rapidly and, in tandem, new visions of how GIS can be exploited for purposes of historical research continue to emerge (Bodenhamer,

Corrigan, & Harris, 2015; Knowles A. K., 2016; Schlichting, 2008). Initiatives such as the Stanford Spatial History Project have appeared as incubators for creating new ways of looking at, and "doing", history, seeking to examine historical events in ways that augment traditional historical research (White, 2010).

Even with this promising outlook, HGIS has not been without its detractors. Pushback from the essentializing nature of the quantitative revolution and structuralist methodologies was expressed in the form of what has become critical GIS (Harris, Bergeron, & Rouse, 2011; Knowles A. K., 2016). Critical geography and critical GIS seek to bring a critical socio-theoretic and cultural perspective to geographical research. This criticality delves into the consciousness of historical actors to derive motivations rather than 'simply' analyzing results (Guelke, 1997). In addition, the role of positionality and reflexivity on the part of the researcher along with consideration of representation are recognized and incorporated into research.

The spatial humanities lie at this intersection of critical geography, GIS, and the Spatial Turn, or as Harris notes "the engagements between GISc and the humanities" (Harris T. M., 2017, p. 2). As Harris et al. (2011, p. 238) points out "The importance of the Spatial Turn in the humanities is to not only build on the methods that are central to geography but to also shift focus toward geographical concepts and spatial thinking in order to gain spatial insights not previously seen". Arising from this mix of concepts is *deep mapping*, which can be seen to mirror the *thick description* of anthropology (Knowles A. K., 2014), and is now coming into its own as a means of expressing these critical ideas within a computer-based context (Bodenhamer, Corrigan, & Harris, 2015). Deep mapping and its related concepts move beyond presenting lines and points on a map, they provide mechanisms that allow viewers to interact, explore, and question events through the use of embedded imagery, documents, video, and links to external resources. By embedding

additional historical sources within a spatial story, deep mapping provides an evocative and deeper view of historical events. For example, in a deep map the routes traveled by people may be color coded to indicate the emotional state of the travelers and to draw out the emotional comparisons between groups. A deep map might also provide interactive options that would link locations on the map to transcripts or images of relevant documents or imagery or the use of Virtual Reality to provide a better sense of place. The point to be made is that by marrying the mapping process with the historical record a more complete story can be allowed to unfold.

Current humanities research is also experiencing an inundation of large volumes of digital data (Lord, Macdonald, Lyon, & Giaretta, 2004). These data and derived data generated as part of the research process are often considered by-products by humanities researchers and may be abandoned at the completion of a project. However, rather than being by-products these data are increasingly seen as artifacts that have value in-and-of themselves. Importantly, the intrinsic value of these data extends beyond the scope of the original research project and beyond a project completion date (Kong, 2016). As with other physical artifacts, the curation, that is the maintenance of long-term accessibility to, and viability of, the data is now viewed as an important aspect of the research planning process. This has led to the evolution of digital data curation, a field that is taking on an ever-increasing role within library and archived sciences.

An important aspect of digital data curation is to provide sustained access to the artifacts of historical research, whether physical or digital. This can be for collaborative purposes as during a research project, or for subsequent use in other research programs. In this respect curation takes advantage of *neogeography* – an end-user oriented and globally accessible means of accessing geospatial data and the collection and analysis of geographical content by a broad community of non-expert users (Harris, Bergeron, & Rouse, 2011; Harris T. M., 2017). Whereas the processing

and storage of digital data was founded on advances in computer hardware, data access and dissemination are based on the growth of the internet and the development of metadata standards for document content tagging, search, retrieval and data interchange.

Addressing the Data Collection and Geospatial Challenges Faced by Historians and the Humanities

An obvious challenge to historians and humanists lies in discovering and gaining access to primary source documents. For example, the marriage and baptismal recorded in early modern times are likely to be found in churches, while notarial records, produced under the jurisdiction of a civil authority, may be located in local municipalities or in the archives of central repositories. But a particular challenge faced by historians is the task of examining and collating these unstructured materials, such as the records at *Sagrada Catedral de San Christobal*. There is the distinct possibility that such physical documents may have deteriorated, pages torn, or individual folios may be separated from their original bindings. In addition, paleographic style, language, hand-writing, and the abbreviations of individual scribes must also be navigated to accurately interpret the author's intent. Heidorn (2008) suggests that in contrast to indexed data stored in formal data repositories and available online, these unacknowledged 'dark data' are essentially invisible to researchers and the public. Yet, it is these data that will likely form the essential information necessary to advance the research frontier in historical studies. Furthermore, while current research is often facilitated by computer-based technologies that streamline access to digital source data and documents, these unstructured primary source historical materials are considerably more problematic and arguably require more sophisticated technological solutions.

Several challenges can be identified when working with these types of primary source materials. One major challenge is the extraction of spatial content from these unstructured historical materials not least because of the often-imprecise spatial content embedded in them. Considerable effort has been applied toward developing methods for the automated parsing and interpretation of geographical content from unstructured texts (Gregory, Cooper, Hardie, & Rayson, 2015; Southall H. , 2003). Natural Language Processing (NLP) techniques have been developed to extract information based on the syntax and semantics of written language, and gazetteers have been used to identify references to known locations and to assign geographic coordinates (Martins, Manguinhas, & Borbinha, 2008). But these processes are far from simple. For example, in the sentence "She lived with her husband on a riverside plantation located thirteen miles south of the city that would later bear his name" there are references to two people, a river, and a city, along with an indication of distance. For an historian, knowledge of the name of any one of these features – Potomac, Martha, George, Washington – derived from the context of the surrounding text - would help augment the record and provide a spatial context for the statement. For computer-based systems, this type of implicit, intuitive, and contextual understanding is difficult to replicate. Encouragingly, progress has been made in the development of technologies and techniques for automating the extraction, management, and analysis of data from historical texts (Alex, Byrne, Grover, & Tobin, 2015; Berman, 2009; Berman, Mostern, & Southall, 2016; Gregory, 2002; Gregory, Donaldson, Murrieta-Flores, & Rayson, 2015; Martins, Manguinhas, & Borbinha, 2008; Martins, Anastácio, & Calado, 2010; Santos, Anastacio, & Martins, 2015; Southall, 2003; Southhall, 2011). With advancements in technology, both applied text processing and text content management have progressed rapidly over recent years resulting in the ability to automate a great deal of the mechanical portion of historical data analysis.

A second challenge related to working with historical documents is that once data has been extracted and organized, historical scholars must then evoke a deeper meaning to the spaces and content (Harris T. , 2015). An understanding must be developed regarding the ways in which social and personal meaning can be conveyed that is richer than the more traditional methods of map making employed within GIS. This has proven to be a non-trivial task. New ways of attaching emotion, experience, and perception through visual representations, for example "inductive visualization" (Knowles, Westerveld, & Strom, 2015), have been proposed, yet truly intuitive and meaningful methods have remained elusive. The question must be asked whether humans have inherent perceptual limitations or biases that will always form a barrier between quantitative representations and emotional response (Drucker, 2011). In addition, emotional response itself is a very personalized concept. The sound of a bagpipe, for example, may be considered either beautiful or grating depending on the listener. There is also the companion issue of making sense of the significance of not only individual information but the interrelationships that might exist between them. The disparate nature of research materials and research results, which may exist in a variety of forms and formats such as scanned images of documents and map images of works of art, and 3D models, makes their analysis as a coherent assemblage difficult to achieve. While these materials provide valuable information in and of themselves, the ability to link them to text, maps, and to one another, for the purpose of analyzing possible relationships, is largely unavailable at the present time.

Third, regardless of the research question or the format of the available materials, historians often need to collaborate with others who may sometimes reside in different geographical locations. This has implications for the collection and recording of data from spatially dispersed entities. Two forms of collaboration can be identified that pose challenges to researchers;

collaboration supporting the collection and assembling of data, and the collaboration required to analyze those data. Collaborative data collection can occur in many forms including crowd-sourcing and volunteered data collection. However, the application of these technologies can go untapped due to the lack of an accessible technical infrastructure that enables such collaborative efforts. Collaborative analysis, which requires seamless interaction, demands fluid communications that support verbal, visual, and sometimes physical interaction. Collaboration, therefore, involves developing interactive capabilities that brings researchers together in the same virtual space even when geographically separated.

Finally, during the research process, historians collect data and produce materials such as notes, transcriptions, translations, and metadata. These valuable materials are often stored informally by the individual researcher or even discarded upon completion of a research project and are therefore unavailable to other scholars – in essence, entering the domain of *Dark Data*. Beyond the publication of data, the preservation of collected and derivative materials is often an afterthought in humanities research (Kennan & Markauskaite, 2015; Lord, Macdonald, Lyon, & Giaretta, 2004). Research projects begin with specific research goals and once satisfied, the long-term preservation of the artifacts of the process, digital or otherwise, may seem to be of little concern to the research team. In addition to the lack of prioritization, curation can appear to be a daunting mass of bewildering standards and protocols, the mastery of which may challenge most researchers in terms of time, cost, or expertise. Making the results of humanities research discoverable and available so that others can benefit from them is critical if they are to be of use for future study. Maintaining viable digital data collections requires the need to understand geo-spatial data management, curation, and interchange technologies and to apply them to physical and digital archival management practices (Harris, Rouse, & Bergeron, 2010).

Individually, each of these research challenges constitutes a significant portion of the data collection and analysis effort, yet the application of off-the-shelf GIS and database software, while automating some of the work, currently does little to provide a cohesive environment to support historical research operating under the challenges identified above. While other projects, such as the *Pleiades* gazetteer of ancient places (Pleiades Individual Contributors, 2018), the *Roman Roads Research Association* gazetteer (Roman Roads Research Association, 2018), and the *Transkribus* project (Digitisation and Digital Preservation group, 2018), offer support for individual challenges, none successfully provide a comprehensive research solution. The premise of this research study is that by integrating the collection of data, the extraction of content, and the analysis of information from what has traditionally been post-data collection analysis and research process, more efficient processing and more effective historical research can be achieved.

The Research Goal

To this end, the research goal of this study is to develop, implement, and demonstrate an integrated customized geospatial system capable of assisting historians in the collaborative real-time capture and management of structured and unstructured data contained in archival source materials and that will facilitate subsequent data mining and geo-historical information exploration. To achieve this, the following research questions are addressed.

1) How might the shortcomings of off-the-shelf GIS be addressed through a customized system embedded within the concepts of the spatial humanities and deep mapping?

2) What system functionality is required by historians to facilitate the transcription, normalization, and translation of historical texts?

3) What is the functionality required to extract and use spatial data held in historical texts and how might the resulting information about individuals be assembled into family genealogies?

4) What database structure is optimal for managing historical content in a manner that supports the integration and linking of people, places, and events?

5) What system mechanisms are required to facilitate the discovery, linking, and embedding of internet resources related to the historical texts, people, places and events captured from primary source materials?

6) How can a software user interface (UI) be optimized to support seamless extraction of historical content and interactive data exploration to streamline and facilitate the work of historians?

7) What system functionality is required to support the export of digital content for ingestion into archival and curated systems?

8) Finally, how can the findings of this study be evaluated in terms of system functionality and benefit to the historian?

The Spatial Historian is the result of an investigation into determining an appropriate technological framework for addressing the challenges facing historians when working with unstructured historical materials. The Spatial Historian offers an integrated view of the interrelationships of people to their cultural and physical environments and supports a better understanding of the role that spatial context plays in human historical behavior. Importantly, the system incorporates scholarly historical research into the technological processing of historical materials thereby embedding the knowledge of the historian within the collection and processing

of historical data and facilitates real-time interpretation and exploration. By using additional technologies such as virtual reality (VR), automated text parsing, and toponym lookup, this research addresses key technical difficulties often encountered by both GIS and historical researchers related to inserting space and spatial analysis into the process of data extraction from unstructured primary source materials. The system enables the capture and management of data contained in the historical archives and facilitates data mining and information exploration within a single comprehensive software environment. In addition, the system supports the temporal analysis of historical information and provides the ability to examine correlations between exogenous factors, such as climate and disease, with observations of events recorded in the local historical materials.

Methodology

In response to the identified limitations of existing geospatial technologies an initial investigation was conducted in the fall of 2015 into the development of a loosely coupled GIS designed to facilitate the processing of historical texts and specifically the early modern texts of St Vincent, in what is now part of the Italian Alps (Schindling, 2016). The system consisted of a basic user interface that was specific to the examination of notarial records from the sixteenth-century town of St. Vincent. This prototype system was intended to insulate the historical research work from the complexities of the underlying technology. Using the system, information could be extracted from the source materials about people and property. Researchers used an interactive 2.5D map interface which enabled them to query and explore the location of family members, and the distribution of land holdings, and land use patterns. While the prototype reinforced the potential need for such a system for furthering the historical understanding of a particular time and place, it

also exposed many system limitations and opportunities for enhancements. For example, it became apparent that researchers would benefit from the ability to include data from more than one study area and to share data across multiple research projects. This concept of shared data requires standards-based web access to both digitized maps and data extracted from the texts. It was also seen that the integration of deep mapping concepts contributes significantly to a richer conceptual understanding of the historical investigation.

Developing such a customized and appropriately framed system along with its supporting computer architecture is clearly critical to the success of this research goal. When developing software solutions one important early question is whether the functionality should be custom developed or whether it should be assembled from preexisting functional components. At one end of the spectrum, fully customized applications allow the system to be tailored and fine-tuned to the specific needs of a project. The downside is that fully custom coded systems require a long-term support commitment and significant technical expertise in a variety of, sometimes seemingly esoteric, computational technologies and database and software platforms. At the other end of the spectrum, development can be based on assembling publicly available open-source software, web services, map services, and data services. While this approach can reduce development time and maintenance concerns it can also lead to a disjointed, '*cobbled together*', system that may support individual functional requirements but does not provide a cohesive polished and shared user experience. This study pursues a customized approach.

The custom system presented here leverages external services for many of the specific processing functions. An example of such an external service is that of speech-to-text conversion. This functionality is a complex process which is compounded when there is a requirement to support multiple languages. Developing this technology requires a deep understanding of

linguistics, sound processing, the ability to adapt speech interpretation to the idiosyncratic speech patterns of individual speakers, and other skillsets. Development and on-going maintenance of this type of highly technical and domain specific functionality is well beyond the capability of any single individual. Fortunately, many functions, such as speech-to-text conversion, are currently available through services provided by large software venders such as Microsoft, Google, and ESRI.

The user interface (UI) is perhaps the most visible portion of software systems and of central importance to the actual users. It is what the system user sees and interacts with, and therefore its design must be intuitive and allow access to the necessary information and functionality. Less well recognized is the technical foundation upon which the UI relies in order to handle data management, user authentication, internet access, and other services. The user interface is supported by a series of information services that provide these foundational functions. Information management is provided by a database tier for storing geographical information and attachments used in map production along with project related details and ancillary data used to support the user interface and manage non-map related system functions. The custom database also supports the storage of Open Geospatial Consortium (OGC) compliant spatial data types using standards for linking various project artifacts and resources to geographical locations.

In the Spatial Historian, information from a database is accessed directly by a higher-level service tier process through a custom developed service process with a custom service Application Programming Interface (API). Separate functions that support adding, editing, querying, and deleting data are thus accessed through the APIs. Both the mapping service and the custom API are designed to support distributed web access. By implementing services in this manner, the architecture allows users at remote locations to access the underlying data in real time over the

internet without requiring direct database access. The layers of abstraction created by the tiered design provide greater flexibility and add additional levels of security to the system.

To facilitate collaborative research, an important design consideration of the Spatial Historian is the need to support mobile devices. This capability allows collaborators in disparate locations to collect and edit data and imagery and attach geographical coordinates to the data. Because all data updates within the system occur in real-time, information collected in the field is immediately made available to all researchers working on the study regardless of their physical location. Mobile access also provides a mechanism for researchers in the field to ground-truth and add additional data by visiting relevant archives and comparing previously recorded information with that which is visible onsite.

Long-term preservation of the data that is collected and generated is ensured through the recording of Metadata Object Description Schema (MODS) compliant descriptive metadata that can be exported in XML or ASCII flat file format for ingest in external archival systems (NISO, 2004; RLG Task Force on Archiving of Digital Information, 1996). All geospatial data in this system are capable of being exported in industry standard formats with accompanying Federal Geographic Data Committee (FGDC) FGDC-STD-001-1998 or International Standards Organization (ISO) 19115 metadata (Bleakly D. R., 2002; Bose & Reitsma, 2005).

Case Studies of St. Vincent and The Slave Societies Digital Archive

The research for this project was reinforced through the use of two case studies, each chosen based on the unique challenges that they presented, the types of available primary source materials, and their differing research objectives. The first case study is based on research

examining ways in which the familial and social structure of the sixteenth-century town of St. Vincent, Italy are reflected in notarial records and other documents recorded during the period. The second case study is based on the work of the Slave Societies Digital Archive, which is dedicated to preserving ecclesiastical and secular documents related to Africans and those of African descent in the slave societies of the Americas. Because the Spatial Historian has been designed as a platform that supports the processing of a variety of primary source materials with potentially differing historical research objectives, the markers for success vary by case. These evaluative metrics include basic system functional capabilities used for all projects, such as the support for storing raw transcriptions and extracted content. In addition, the achievement of 'success' for each case study must be defined individually based on the specific objectives of the research team. The two case studies selected are distinct projects in their composition, in the source materials available, and in their objectives.

The case study of the sixteenth-century town of St. Vincent, Italy, is based on research performed by Dr. Matthew Vester of West Virginia University. The available text materials used in the study of St. Vincent offer examples of documents representative of society at that time period and place. The materials comprise notarial records from the *Archivio Notarile Distrettuale di Aosta* for the city of St. Vincent, now Italy, documenting, among other events, the transfer of property and land between individuals. The notarial records of St. Vincent contain an indication of where the transaction or event were undertaken along with the names of the parties involved and the individuals present at the recording. In addition, the records provide explicit details about the location and size of the property being transferred, and the relationships of the parties involved.

Another important element of the St. Vincent case study is the availability of hand drawn maps of the region that were created from a survey conducted the 19th century under the

commission of the Italian government. These maps offer an exercise in the techniques required to convert historical maps into digital formats amenable for use in a GIS. Integration of these maps also allows the system to be evaluated based on its ability to link the people and properties listed in the notarial record to their geographical locations.

The Slave Societies Digital Archive study is dedicated to preserving ecclesiastical and secular documents related to Africans and those of African descent in the slave societies of the Americas. The project, directed by Dr. Jane Landers at Vanderbilt University, was initiated in 2003 with funding from the National Endowment for the Humanities. Initial work focused on capturing the digital images of endangered historical documents with the objective of preserving their content and making them available for subsequent research. The primary directive of the Archive is to "preserve and freely disseminate invaluable documents for the history of Africans and their descendants in the Atlantic World" (Vanderbilt University College of Arts and Science, Department of History, 2018). The materials in this archive are comprised primarily of records of baptisms, marriages, and deaths, and are of a distinctly different structure and content from the records of St. Vincent. The archive currently contains over 310,000 document images from records located in Brazil, Colombia, Cuba, and the United States, and with references to an estimated four to six million individuals.

CHAPTER 2 - Space, GIS and Historical Investigation

The Spatial Turn

The Spatial Turn is now an integral component of humanities research. The recognition that space plays a vital role in human behavior harks back to the 1960s. In the words of Ethington (2007, p. 478):

> Spatiality as an analytic category finally came into its own in the 1960s and 1970s, with the work of Gaston Bachelard [1958], Michel Foucault [1967], Henri Lefebvre (1974), and Yi-Fu Tuan (1977). Thanks primarily to the work of Edward Soja (1989, 1996) and David Harvey (1989), a 'spatial turn' has occurred in many of the human sciences.

Central to the spatial turn are the spatial connections embedded within social contexts and the social spatiality that individuals intuitively construct in their tier relationships one another. In these encounters the closeness and distance in the relationships of individuals to others, arise not only in a Euclidian and geometric sense but in an inter-personal sense. Social distance and associated social manifestations are mediated by, and judged from, within an individuals' and societies cultural context.

A leading voice in articulating the influence of space on society was Lefebvre (1991) with his publication of *Production de l'Espace*. The influence of Lefevre is summed up by Ethington when he writes:

> So successful have Soja and Harvey been in spreading the pathbreaking ideas of Lefebvre that it is no exaggeration to call the current discourse on spatiality 'Lefebvrian' (Ethington, 2007, p. 478)

In *Production de l'Espace,* Lefebvre presents his often-cited triumvirate classes of spatial understanding that include *Spatial Practice,* which describes the use of space in the context of everyday life and how space is navigated and used. Second, Lefebvre introduces R*epresentations of Space*, or how space is consciously represented in culture through architectural place, city plans, and other spatial designs. And, finally the *Spaces of representation* includes the literature, art, and structures that are the culturally significant manifestations of our existence.

These spatial classes provide a comprehensive framework within which space in its multitude of social contexts can be evaluated. Yet Lefebvre was very much the product of his politically active life as an active leader in the French communist movement and as a leading member of the French Resistance during World Word II (Soja, 1989). Lefebvre himself was critical of the proliferation of *spaces* in literature suggesting that their multiplicity entailed that related *turns* must be questioned as well (Lefebvre, 1991). For Lefebvre, *space* is a purely political construct; an arena in which the battle between capitalism and communism is fought. In cases where the political link is tenuous, Lefebvre insists on the existence of hegemonic capitalist/neo-capitalist forces imposing their own control of space in a never-ending class struggle. Lefebvre, in essence, creates his own specialized version of space that is purely political in nature and one that is consciously constructed by unseen/omnipotent (largely capitalist) institutional entities. While he purports to demonstrate that space and the production of space have an abstract universality, he nonetheless limits his scope from a truly universal one without the involvement of humans to a purely anthropocentric one that is grounded in politics without ever seeming to notice the contradiction. Lefebvre also acknowledges that certain aspects of his ideas may not apply to different cultures, specifically pointing out the distinction between his representational space may not apply in the "East", and specifically China. Lefebvre (1991, p.48) focuses on the capitalist and

neo-capitalist production of abstracted space whereby 'natural' spaces are transformed to political spaces by means of symbolic mediations such as statues, monuments, temples, and other edifices.

The writing of Yi-Fu Tuan provides a less politically charged version of space. As a humanistic geographer, Tuan sought "understanding of the human world by studying people's relations with nature…" where space is epistemologically situated within the context of human nature and the environment (Tuan, 1976, p. 266). His application of ethological studies is in stark contrast to the politically motivated vision of human action implied by Lefebvre and Harvey. Yet Tuan builds his arguments on an understanding of non-human animal societies and on broad generalizations about human personality characteristics (Tuan, Space and Place: The Perspective of Experience, 1977). Tuan's broad generalizations such as "Mathematicians doodle" (Tuan, 1975, p. 209) and that geographers have the "greatest tendency" (p. 210) to create and arrange what he calls mental maps and perhaps faults in his logic but are essentially irrelevant in the construction of his arguments even though they do indicate that he is prone to apply preconceived and unsubstantiated ideas to rationalize his theses.

David Harvey is one of the quintessential converts from the reductionist/masculinist quantitative geographer to the world of critical/Marxist geography. So important has been his role that Castree (2005, p. 62) states:

> From the early 1970s the idea that geography could be a spatial science came under attack, not only by those who were never fans in the first place but also by several of its former advocates. …Chief among them was David Harvey, whose Social Justice and the City (1973) paved the way for Marxist geography.

Early in his career, Harvey wrote a paper that attempted to develop a scientific theory of human behavior that governed geography (Harvey D. W., 1967). At that time he concluded that with the development of more advanced behavioral models, it would be possible to model

"aggregative and general" geographical theory; one that would allow social behavior to be predicted and manipulated through the use of computational algorithms. By the early 1970s, Harvey approach to such modeling and the quantitative revolution had emphatically been abandoned (Harvey D. W., 1974). During this transitional period, he developed a distinctly Marxist mindset, taking on the character of a political commentator and anthropologist. In the title of his 1974 paper he asked, "What Kind of Geography for What Kind of Public Policy?" (Harvey D. W., 1974). What he was asking, however, was not which methodological approach should be applied, but rather what should drive geographical research and to what end should it be targeted. Harvey questioned whether geographical research should be directed by the desire to publish, the ability to find funding, or by higher moral standards that rise above personal gains or capitalist/institutional objectives – whether the motivation for research should be altruistic or whether it should be driven by personal, corporate, or state centric motivations.

In 1984 Harvey followed this paper with *On the History and Present Condition of Geography: An Historical Materialist Manifesto* (Harvey D. W., 1984). In this paper he presents a clear discussion of the crisis of identity within geography since the quantitative revolution. This point is reiterated in his 1990 paper *Between Space and Time: Reflections on the Geographical Imaginatio*n (Harvey D. , Between Space and Time: Reflections on the Geographical Imagination, 1990). In this paper Harvey suggested that as geographers broke free from the positivist mold, they failed to create a cohesive voice but instead became "a veritable cacophony of competing messages" (Harvey D. , 2001, p. 116). *Place*, according to Harvey had recently become the leitmotif "used to provide the discipline with a more powerful (and equally fictitious) sense of identity…" (Harvey D. W., 1984, p. 4). As with his other work, Harvey was presenting a political agenda for the field based on an inwardly looking critique and presented reasons why his preferred

Marxist philosophy, or an alternative system of anarchy, would not gain widespread political traction. Importantly for this study, while he presented a manifesto for geography, Harvey perhaps did so with surprising naivete as to the possibility of success for he failed to apply the same critical thought to the potential for success of his manifesto as he applied to Marxist politics. Harvey argued that geography must be a "peoples' geography" (Harvey D. W., 1984, p. 7), though lacking from his conclusion in his 1984 manifesto is a concrete a plan of action – the means to transition from theory to practical application.

The unstated result that comes through in Harvey's migration from the quantitative geography of the 1960s to the Marxist geographer of the 1970's is apparent in the distance he places, consciously or sub-consciously, between himself and the application of his early theoretical models. His rejection of quantitative research may be interpreted as a rejection of its validity, but it is apparent from his subsequent writing that his change of direction was due more to a shift in his moral conscience rather than a lack of faith in the viability of quantitative research. Harvey is viewed as a leader of Marxist/critical geography but by abdicating his leadership in quantitative social modelling he, in effect, handed the keys to the castle to the very capitalist institutions that he most adamantly denounced. The advanced behavioral models that Harvey saw as lacking in the 1960s have clearly been developed in the ensuing years as evidenced in recent years through their application by corporations such as Google, Facebook, Amazon and others (Zuboff, 2019). Indeed, in what is the most significant manifestation of the potential of technology to date, the Russian government successfully manipulated public opinion through the use of social media to the extent that the results of the U.S. presidential election have come into question (Mueller III, 2019).

Space and Technology

Social theoretic discussions address the *what* and, to an extent, the *why* behind the awakening of a conscious consideration of space as a vital agent in human society. The *how* aspect of the spatial turn is addressed more directly by technological innovation. Bodenhamer (2015), suggests that there have been two technology driven spatial turns, the first beginning in the 1880s, corresponding to advances in transportation and communication, and running through the early 20th century, and the second beginning in the late twentieth century. Both turns were initiated by technologies that changed the perception of distance through advances in the speed at which information, goods, and individuals could be. The technology driven spatial turn of the 1980's is defined by the application of GIS to humanities research. Guldi (2017) describes this as "an impulse to position these new tools against old questions", and Bodenhamer (2015) describes it as "a re-emergence of space and place as important concepts in the social sciences, driven in large part by GIS and other spatial technologies". Bodenhamer also points out that there have been different variants of spatial turns within the humanities corresponding to differing research areas, occurring at different times and with different pedigrees. For the purposes of this paper, the primary concern is the spatial turn as applied to historical research.

Conceptual Challenges

GIS as a technology was developed in the mid-1960s ostensibly as a means of producing digital maps (Goodchild, GIScience, Geography, Form, and Process, 2004; Knowles A. K., 2000). Initially, these software packages were home-grown and had few analytical capabilities. Government and resource and utility management demands contributed to the demand for and

growth of GIS and, more recently, concerns over rapid changes to the global environment continue to drive rapid advances in the data management, representation, and analytical capabilities of GIS. With these advances, large volumes of digital spatial information can be captured and analyzed. New avenues of GIS applied research within the humanities have drawn upon the technology building heavily on the availability of tabular data such as historical census tables and reports. GIS is most readily associated with map making for "Maps serve as one of the most important repositories of both the raw data and the results of geographic inquiry, and mapmaking has always figured prominently in the skill set of geographers or their supporting staff. Maps are thus important and indispensable tools in the geographer's search for understanding of how human and physical processes act and interact on the Earth's surface: of how the world *works*." (Goodchild, 2004, p. 709) GIS has also been termed a visual language as well as a spatial language (Knowles A. K., 2014) and offers a means of performing spatial analysis that is not available through other means. (Knowles A. K., 2000). GIS emphasizes the role of space in human affairs and, in the context of this dissertation, can reveal potential spatial linkages between historical events and human actors that are not easily perceived or derived from the aspatial examination of historical documents alone. In other respects, GIS could be regarded more as a database technology than a mapping technology (Gregory & Healey, 2007). By applying the expressive visualization of GIS along with its information management capabilities, GIS is capable of complementing humanistic research to produce a comprehensive understanding of human activity – past and present.

Because of the capabilities of GIS, it is uniquely positioned as a tool to support historical research (Knowles & Hillier, 2008). The capacity of GIS to store, link, and retrieve information in spatial and temporal contexts opens new avenues for analysis and allows researchers to ask new sets of questions (Bodenhamer D. J., 2008; Goodchild, 2008; Gregory & Healey, 2007). Beyond

being a mapping technology, GIS, used in conjunction with recent advances in spatially enabled multi-media, geovisualization, and web services, offers a powerful tool for the development of spatially aware historical knowledge. Thus, acknowledging that location is an important part of historical narrative (Jessop, 2004), Historical GIS (HGIS) arose from the application of GIS in historical studies and made space a central component of historical analysis (Knowles A. K., 2000). HGIS, combines numerical information, such as census data and other digital geographical information to produce a variety of location-based visual representations (Gregory & Healey, 2007). Historical GIS is predicated on accessibility to substantive computing power and digital mapping that has only reached a degree of maturity in GIS within the past decade. Even now GIS capability continues to develop rapidly and, in tandem, new visions of how it can be exploited for purposes of historical research continue to arise (Knowles A. K., 2005; Schlichting, 2008).

Historical Geography originates from a different epistemological view of the study of history that explicitly recognizes space and geography in human affairs. With its origins within geography, historical geography recognizes the effect of human activity in shaping spaces and in the human production of place through both physical action and the cultural assignment of significance. The related and overlapping term 'spatial history' has similarly been used in acknowledgment of a long desire to integrate historical and geographical scholarship (Pavlovskaya, 2006). Recent advances in GIS capability has changed the calculus by enabling space to be integrated into historical narrative in a way that was not previously possible. Bodenhamer (2010) contends that "GIS lies at the heart of the so-called spatial turn" and that the spatial turn may not be a watershed event so much as a realization of long-standing desires to examine the spatial context in which historical events are played out. This realization has been heavily influenced through increasing use of GIS (Harris, Rouse, & Bergeron, 2010).

In more recent years access to GIS products such as maps have greatly increased with the availability of web accessible geographical information layers such as topographic layers, administrative boundary layers, and road layers (Gregory & Ell, 2007). This in turn has aided the growth and utility of HGIS. In addition, with these layers there are highly accurate elevation data that are publicly available from commercial and governmental sites such as the USGS, NOAA, ESRI, Microsoft, and GOOGLE. These data, in combination with current GIS and mapping software, have opened up the field of Historical GIS and made it more accessible than what could have been imagined only a decade ago. Yet, Knowles comments "that virtually no one working in HGIS has demonstrated how to link relevant theory to GIS methods and analysis" (Knowles A. K., 2014, p. 209) and that applying GIS to humanistic scholarship is a significant challenge. Given Knowles' description of GIS as a spatial language, this lack of theory may perhaps be expected. Written language, while based on defined vocabularies and syntax, does not prescribe a formula as to how best to convey information or emotion. Similarly, the message conveyed by any given HGIS representation is based on the historical context, the objectives of the historian, and the authors framing of the output. Due to the subjective, interpretive and nuanced nature of historical scholarship and the disparate nature of primary and secondary source materials, it is quite unlikely that a single theoretical underpinning to the application of GIS for historical research is possible. Historical research requires a reading of the past that is seen from the perspective of the actors after accounting for the filter applied by the chroniclers and the cultural biases of the time. Historiography is the history of historical interpretation, the variety of historical methods and approaches, the ways in which historical practice relates to social and critical theory, and the histories of interpretation in specific historical subfields (Vester, Personal Interview, 2014). In this

sense, HGIS is a method for historical research, an aid to historical interpretation, and a means of representation.

Discussion to date downplays the advances in the theory of spatial information management and of geoscientific thought that has been applied to spatial analytics and GIS. In an effort to take GIS beyond an uncritical focus on map production and its characterization as a "non-intellectual expertise" involving little more that the pushing of buttons, Goodchild proposed the field of GIS Science (GISc) (Goodchild, 1992). GISc was essentially the integration of spatial analysis with GIS methods, technology, and data (Goodchild, 2004; Goodchild, 2006; Jordan, 1990) and a means to better understand the processes and spatial relationships underlying standard GIS map representations. GISc sought to advance both the technological underpinnings of GIS and the application of GIS within the context of a theoretically informed scientific and humanistic framework (Goodchild, 2006). Advances in technology have made it feasible to implement the geographical concepts underpinning GISc such as the application of Tobler's first law of geography in spatial autocorrelation and spatial regression (Tobler, 1970).

As with other aspects of the *quantitative revolution*, there has been hesitancy in adopting GIS within the humanities because of the perception, if not reality, that the technology was positivist, reductionist, and implied a sense of authority with definitive answers. Bodenhamer, an historian and one of the leading advocates of spatial technologies, contends that the technological underpinnings of GIS are fundamentally unsuited to the ambiguity, unstructured, and non-linear nature of humanities research (Bodenhamer D. J., 2008). GIS, he argued, portrays the behavior of individuals via the structure and logic of computer-based systems poorly represents human behavior (Castree, 2005). However, in countering this idea of essentially random human actions, "critical realism" presents a philosophy that is more amenable to developing models of human

activity (Castree, 2005) whereby human action can be observed over time and ordered patterns discerned. The observed patterns can be analyzed in context to develop an understanding of the factors that condition them and the underlying processes that influenced their development. The often-cited conflict between the quantitative nature of GIS technology and the subtle, nuanced interpretation required to develop rich historical insight is in fact as much a division within geography as it is between geographers and historians (Viles, 2005). And, while these characterizations are credible, the aggressive argument against GIS as a masculinist/capitalist tool could as easily be applied to written narrative or other forms of media. However, GIS, possibly because of the relative newness of its underlying computer-based technology, became a lightning rod for postmodernist criticism (Pavlovskaya, 2006).

The sense of authority and objectivity that computers and maps imply in the representation of reality while generating a sense of value neutral 'answers' tends to separate the viewer from the underlying nuance of social and personal reality (Harris T. M., 2017). This detachment manifests itself in modern life where relying on a GPS to navigate can aid in arriving at a location but at the same time detracts from the ability to learn one's surroundings or see the places that were passed on your way to a destination. A traveler may get from point to point yet the connection to the surroundings is lost. Similarly, maps can show a set of features or a path but can also lead one to believe that what is not represented does not exist.

GIS has also been construed as being an 'answer machine' and no more. Picasso's comment that computers "…are useless. They can only give you answers." (Fifield, 1964) was a view held by many researchers in the humanities, yet it arguably misses the point while hiding the real potential of GIS; that is, rather than being a producer of answers, the true value of GIS is its ability to expose new possibilities to develop new questions that may have otherwise gone

unrecognized. In the words of Bodenhamer, "GIS can be seen as a technology and/or as a means for advancing postmodern scholarship that embraces multiplicity, simultaneity, complexity, and subjectivity." (Bodenhamer D. J., 2010). By considering space as an historical actor in its own right, a fresh paradigm – the "spatial turn" – could be realized and could add an additional dimension to historical scholarship.

The quantitative / qualitative divide, pitting the positivist imposition of science against post-structural reflexivity is in most respects a false dichotomy, reified by ongoing epistemological misunderstandings and misrepresentations. One effect of these criticisms was to provide fodder for epistemological debates while doing little to advance humanities scholarship. However, by re-conceptualizing GIS as a method of investigation rather than a quantitative technology GIS can be seen instead as a means of opening new spatial avenues of investigation. Pavlovskaya (Pavlovskaya, 2006) argues that the quantitative/structuralist labels that have been assigned to GIS are misplaced and that the assumption that traditional humanities research may actually apply quantitative processes under the guise of qualitative epistemologies. Qualitative analysis routinely employs quantitative methods in scenarios such as the compilation of questionnaire responses and the analysis of focus group results. On the opposite side of the discussion, GIS seldom performs true quantitative analysis and for the most part GIS data represent primarily qualitative information and often employ data samples derived using qualitative selection methods. Pavlovskaya suggests that, rather than looking at GIS in the context of a quantitative/qualitative epistemological debate, it should be looked at as a means of opening new avenues of critical analysis. This development of a *Critical GIS* that looks at relationships and processes, meshes well with the growth of mixed-method research that is based on a continuum of quantitative and qualitative approaches in humanities research (Harris, Rouse, & Bergeron, 2010; Pavlovskaya, 2006).

It is also incorrect to think that qualitative description is somehow singularly valid and that it conveys information that is somehow superior in its content to that of a quantitative representation; the same event can have different qualitative meaning to two different people who experience it as it occurred conditioned on personal background, physical point of view, relationships to others affected by the event. Conveying the sense of an event would require that the perceived sense of each participant is also conveyed. Because of these factors, no one qualitative representation of any event can be considered by a given individual as the reality. The conflation of data with a strictly quantitative analysis and the question of whether GIS is a resurgence of the quantitative revolution are misplaced. Just as mixed-method approaches have blurred the lines between quantitative and qualitative techniques, so too might GIS provide a valuable supplemental narrative to historical research and support scholarship founded in an understanding of the role of space in human activity. Key to the adoption of GIS by historians will be the ability of GIS, or more accurately Geographical Information Science (GISc), to deliver on its promise of addressing humanistic needs and addressing the perceived epistemological disconnect. The term *Geographically Integrated History* has been forwarded to describe the way in which GIS might be used to support the inclusion of space in historical research (Ethington, 2007; Owens, 2007). Complex, non-linear, and ambiguous are all terms used to describe the subject matter of historical research which are at the same time exemplars of the types of unstructured information that do not lend themselves easily to computer-based analysis. Owens (2007) contends to the contrary that the complexity of systems that are the subject of historical research and the large number of variables and volume of data that historians need to manage and synthesize is a strong argument for the inherent data handling capabilities of GIS.

Technological Challenges

Despite much rhetoric, information systems are not eschewed by historians. It can be safely said that just because historians study the past does not mean that they live in it. Historians, and other humanists, are keenly aware of the benefits of applying digital technology to their research and many historical conferences now have panels specifically focused on digital humanities and specifically on GIS. Three of the leading proponents of applying GIS to historical research, Ian Gregory at the University of Lancaster, Phillip Ethington at University of Southern California, and David Bodenhamer at Indiana University – Purdue University Indianapolis, are themselves historians. The growth of the digital humanities has led many researchers and institutional organizations to respond with activities specifically focusing on its advancement, with digital humanities 'Centers' now existing at many universities. A small sampling of initiatives includes the Virtual Center for the Spatial Humanities, a consortium of WVU, Florida State, University of Richmond and IUPUI (The Polis Center, 2017), the Vanderbilt Center for Digital Humanities (Vanderbilt College of Arts and Sciences, 2019), and the Edinburgh Language Technology Group (LTG) (Alex, Byrne, Grover, & Tobin, 2015). Researchers working in the digital humanities include Yaser Bishr whose work evaluated the use of Geospatial Semantic Web markup capabilities as a foundation for application development and information access (Bishr, 2008); Ian Gregory who applied the Great Britain Historical GIS as a product and testbed for evaluating geoparsing, error handling, ambiguity, change-over-time, and data management (Gregory I. , 2002; Gregory & Ell, 2007; Gregory, Cooper, Hardie, & Rayson, 2015; Gregory, Donaldson, Murrieta-Flores, & Rayson, 2015); Trevor Harris who explored the use of immersive technology integrating 3D models, audio, and smell for conveying a sense of place (Harris T. M., Deep mapping and sensual immersive geographies, 2017); Bruno Martins, who investigated advanced techniques for

using gazetteers to extract geo-temporal content from text (Martins, Manguinhas, & Borbinha, Extracting and exploring the geo-temporal semantics of textual resources, 2008); and Humphrey Southall who developed techniques for extracting semantic content from historical narratives and evaluating text markup systems such as those defined by the Text Encoding Initiative (TEI) (Southall H. , 2003; Southall H. , 2012; TEI, 2018).

As Jane Landers stated in a recent discussion one difficulty in implementing technology within historical studies lies in the availability of computer resources, with "…the IT department is geared toward STEM (Science, Technology, Engineering, Math); they are not set up to support the needs of the humanist" (Landers, 2019). Humanists recognize the benefits that computer technology offers, and they may even have a desire to apply it, but what they may not have is the background and training to implement or direct the implementation of the types of systems that can best support their needs (Knowles A. K., 2000). Systems designed for STEM applications often must be 'answer machines' in that they provide unequivocal results to engineers who are designing bridges or optimal airflow characteristics. On the other hand, the systems required by humanists must be designed to support the exploratory analysis of structured, semi-structured, or unstructured data. These systems must allow the researcher to interactively view the data in a variety of ways to enable the examination of data within the qualitative context of the historical setting (Burdick, Drucker, Lunenfeld, Presner, & Schnapp, 2012; Harris, Bergeron, & Rouse, 2011). These requirements lie at the heart of the digital humanities and computer-based support for humanistic research combined with the application of critical humanistic thought.

The use of GIS for historical research is necessary premised on the availability of digital data that is tabular, quantifiable, complete and consistent. GIS data are made up of two components; locational information and attribute data. Attribute data identifies the relevant

characteristics about the location such as the name, population, or agricultural production capacity. In a GIS, this attribute information is represented as rows and columns of tabular data with each column containing a specific type of information in a specific format. This structured tabular format is an essential characteristic of the way in which data are stored in most current HGIS (Gregory & Ell, 2007). For most social science researchers, neatly structured and spatially precise tabular data offers an ideal information format. Rows and columns of well labeled and consistently measured values are ready-made for analysis. However, the records of historical events and phenomena are not always so accommodating in providing information in this way and are often recorded as unstructured script in the form of ledgers and narratives. Names, numerical values, and relationships are embedded in narrative descriptions of events and places. In addition to extracting data from these sources, issues of interpretation are compounded by nuance, changing names over time, and even the use of obsolete units of measurement.

Historical sources are often fragmented, sparse, imprecise, and ambiguous. In addition, spatial references are often colloquial and relativistic with spatial extents changing over time. Temporal references are often based on inexact or obsolete reference systems. Historical materials also often suffer from temporal lacunas in the data in which data available for one period of time may be followed by a period of time with no references in the historical record. In these situations, it may be that the missing data are not relevant and can be ignored. However, when this is not the case, some form of temporal interpolation may be required (Dragicevic, Marceau, & Marois, 2001; Gregory & Ell, 2007). These types of '*messy*' data do not fit within a standard tabular data model that is standard for a traditional GIS. The extraction of even this '*messy*' data from primary source materials poses significant technical issues. According to Bodenhamer (2008), "One of the most cited impediments is the technology's awkwardness or inability in managing ambiguous,

incomplete, contradictory, and missing data". Historical texts were recorded for reasons that were relevant in their original context and the rigor and consistency that was applied to the historical record can be highly variable. Inconsistencies in spellings and informalities used when referencing people and places can lead to ambiguity and uncertainty, both of which make the use of GIS problematic (Gregory, Donaldson, Murrieta-Flores, & Rayson, 2015). The texts or, in the case of many modern sources such as Twitter the data streams, are often unstructured and un-mediated. A rule of thumb has been offered among GIS users that as much as 95% of the time spent on any given project is spent preparing the data and the system itself with only 5% devoted to actually producing results (Knowles A. K., 2000). The first challenge in preparing these materials is to identify items of importance such as place names, dates, and names of individuals. Then, once identified they must be validated, disambiguated, and interpreted. For example, the word 'York' may be a person's name (Robert York), part of a title (Duke of York), a place (county of York), part of a place name (New York), or a mis-transcribed name (York shire"). In response, there are many recent initiatives aimed at addressing the 'messiness' of historical data within the context of GIS that range from analysis of text using Natural Language Processing (NLP) to the application of gazetteers to aid in assigning geographical locations to people, place names and events (Manguinhas, Martins, Borbinha, & Siabato, 2009; Santos, Anastacio, & Martins, 2015; Smith & Crane, 2001).

Work is progressing to address data extraction issues and the identification of individuals (or other named entities) within texts. It is sometimes possible to use genealogical services, such as FamilySearch, or other published name lists to help identify individuals, however, except in the case of prominent historical characters, disambiguation is problematic. Models have been advanced for dealing with these issues though as yet there is no commonly accepted, standards-

based, approach that has come to the forefront and managing these data within a GIS is itself a target of research. Gregory (2002) discusses a variety of approaches for handling variation in the spatial definition of named places within a GIS database and offers one possible model. Other alternatives, including data interpolation (Dragicevic, Marceau, & Marois, 2001) to deal with missing data. The development of network models to track spatial changes proposed by Berman (2009), also shows promise. Similarly, Gregory et al. (2015), Martins et al. (2008), and others discuss techniques for extracting and managing these data. Most of these information extraction techniques rely on searching pre-built lists, such as gazetteers and name lists, or performing rudimentary identification of nouns and verbs within the texts. It is hopeful that in the near future more advanced theoretical linguistics will be available for text interpretation. In the commercial sector, Microsoft and Google translators already implement advanced linguistic logic and they both make application programming interfaces (APIs) available that may find a place in HGIS (Google, 2019; Microsoft, 2018).

An additional issue when dealing with large volumes of primary source materials is deciding which texts to read. The traditional approach for historical research is *close reading*, that is historical text are read in depth with the goal of understanding all the nuanced and implied information carried within the words. For a large corpus this close reading of all available materials is not feasible for there is simply not enough time to read every document in detail. To aid in the research of large sets of documents, Franco Maretti proposed the idea of *distant reading* (Alves & Queiroz, 2015; Gregory, Donaldson, Murrieta-Flores, & Rayson, 2015). This idea falls within the context of computational linguistics and consequently is somewhat anathema to traditional historical research. The idea behind *distant reading* is to develop programs that identify significant word patterns and relationships within a large corpus. These resulting patterns can then be used to

identify documents with content, syntactic, or semantic similarities that would otherwise go undetected. Documents that match specified criteria could then be read *in depth* in a traditional manner to determine their significance or their contribution to the subject of study. As with any automated process however, it is important to understand that *distance reading* does not provide definitive any answer but rather suggested reading lists. In effect, *distant reading* is a means of identifying the low hanging fruit that are ripe for investigation and more detailed examination.

The Challenge of Time

An important part of historical research is the ability to track change over time; the ability to trace the sequence of historical, and the propagation of effects and state transitions – both physical and cultural. Yet, in GIS, the issue of handling time, or more accurately, change over time, is a recurring technical problem that is exacerbated when handling historical source materials, which are often analog, fragmented and discontinuous (Gregory & Ell, 2007). In the case of many historical materials, the useful pieces of information that indicate the locations of events or the movement of individuals are embedded in texts, with the texts themselves possibly created by different sources and at different times. One of the complexities of managing of change over time is that the spatial extent as well as the naming of locations and spaces can alter over time. An example of how this can manifest itself can be seen in the changes in definitions of administrative boundaries population changes or counties may merge, leaving some names obsolete and the spatial footprint referenced by the surviving names significantly altered. Or, the name of a place may change due to an extreme event such as military conquest or the renaming of a place in honor of a local personage. Colloquial names may appear and vanish for reasons that are now lost. For example, the description of sixteenth-century St. Vincent as being "contained

within the administrative jurisdiction of the Chatillon family as part of the composite Duchy of Savoy" has no meaning within the current set of national administrative bounding polygons. The village itself still exists as a named entity, although with a larger spatial extent, yet its administrative context has changed completely.

Gregory (2002) identifies the following four types of changes that can occur to administrative units over time: *Transfers*, where part of a unit is transferred from one unit to another; *Name changes* where a unit changes name but its boundaries are unaffected; *Mergers* where a unit is abolished and its component parts are allocated to another unit or units; and *Divisions* where a new unit is created from parts of an existing unit or units. Addressing these complexities of spatial change over time is problematic for most GIS as GISs are based on definable points, lines and polygons. Because of this, techniques must be developed and supporting data structures must be defined to record, track and represent the transformations as the spatial extents of named features change over time (Gregory I. , Time-variant GIS databases of changing historical administrative boundaries: A European comparison, 2002). These techniques include recording *time-slice snapshots* in which individual records are created that fully define an area at specific points in time; creating a *base state with amendments* in which an initial state of an entity along with timestamped changes that are used to construct an area at a specified time; creating *space-time composites* where a base state is recorded together with amendments to fully define an area at a specified time; *date stamping* in which all the individual components of features are recorded along with time encoded attributes; and the use of *transaction logs* where the base state of an area is recorded along with a history of events, or changes, that have effected it over time.

Taking a different tack, Berman (2009) proposed the recording of named places in a network model independent of positional information. These named places can then be associated

with, or disassociated from, different spatial extents at various times. Berman's model also allows a single named entity to be associated with more than one extent at a given time or likewise a single polygon may be referred to by multiple names. An advantage of this model is that it can support many types of spatial change and situations where the same geographical space is assigned to more than one named unit at the same time. This model can also support situations where multiple map sources from different political sources coexist with all variants being readily searchable for overlap and linkages.

To further confound the temporal issue, are debates over the very definition of "time". A compelling argument made by Ethington suggests that time, in reality, does not exist (Ethington, 2007). He asserts that "Most commonly, historians simply confuse time with chronology and chronometry—the 'time' of calendars and clocks" (p. 470) and that what is commonly referred to as "time" is in fact a moniker used to describe the constant reorganization or dynamics of all the bits that make up the universe, Time "…is nothing in itself, but rather a culturally specific reading of the dynamic environment." (p. 471). As such, 'yesterday' 'is' just a different arrangement of what 'is' now. While these debates may appear somewhat academic, they warrant consideration as stimuli for the development of new models of representation. As Ethington (2007, p. 476) contends, the "metaphoric entanglement with space is filled with powerful clues as to the nature of time, and hold profound importance for the debate about historical knowledge of the past".

The Challenge of Representing Time

In visualizing time and change over time, practice tends to fall behind theory. There are theoretical underpinnings in the cognitive sciences that support the idea that visual representations, including maps, provide an intuitive and readily understandable format for transmitting

information (Knowles A. K., 2000). However, the development of intuitive representations of time and *place*, are still under investigation. Knowles discusses many of these issues along with items to consider for advancing HGIS and HGIS, yet she offers no strawman. Even in a recent article she refers to her concept of 'Inductive Visualization' as a *methodology* – possibly conceding that true *theories* in HGIS may not be on the immediate horizon.

Techniques that rely on the manipulation of a static map to imply motion are, as a rule, less than satisfying and largely un-intuitive. The interpretation of static maps requires active thought rather than a perception of change natural based on motion. A more satisfying and intuitive approach involves a cinematic representation where motion – even of abstract polygons – aids perception and comprehension (Ayers, 2010). This is because the eye is adept at detecting change whereas it has a harder time picking out noteworthy features in static images. In the *Where's Waldo* cartoon puzzles the challenge is to find Waldo (the noteworthy feature) within a large jumble of characters. This can be a daunting and time-consuming task, which in the case of a puzzle is the objective. However, for the researcher attempting to identify significance within a large corpus of historical text, "*daunting*" and "*time-consuming*" are not adjectives that are valued. Yet, if Waldo were to move a bit he would immediately be recognized. By applying motion to the data, that is movement over time, noteworthy information may become discernable where in a static presentation it may remain hidden. (Ayers, 2010)

To develop this point of motion further, the winning illusion in the 2017 *Optical Illusion of the Year* contest is built on the notion that motion exposes forms that cannot be detected in a static image (Illusion of the Year, 2017). The illusion starts with a static image that may best be described as a visual version of "white noise"; random black and white shapes with no organization or pattern. Portions of the image or different parts are animated in differing ways exposing obvious

and easily discernable shapes. The key element for this discussion of GIS and the depiction of change-over-time is that once the motion is stopped the once obvious shapes disappear once again into the background noise. Similarly, patterns of interaction and spatio-temporal relationships are difficult to discern when forced to "stand still" in traditional cartographic representations. The technique of expressing change-over-time as movement and the juxtaposition of movements, dovetails with Ethington's argument that "time" is not a thing in and of itself, but is rather a measurement of the change of state and location of all of the parts that make up the universe (Ethington, 2007). In the case of the illusion, the motion exposes objects that "exist" for a measurable "time", yet when motion stops their existence ceases and "time" becomes irrelevant. The reconceptualization of time as motion may free researchers from the burden of attempting to capture other forms of "time" and allow them to look toward more fruitful expressions of the everchanging relationships between events, objects, and spaces.

Other attempts have been made to express a sense of time using maps such as Hagerstrand's space-time maps (Hägerstraand, 1970). These efforts are, however, generally less than satisfying and often require a good deal of effort on the part of the viewer to interpret nor do they provide an intuitive representation that facilitates an understanding of action and motion. As with Waldo, movement is the essential element that provides recognition of that which may be hiding in plain sight. For the Spatial Historian, this idea of motion implies that time-based mapping is best handled with representations that 'move' and interactive representations that allow the viewer to navigate spatially and temporally, (to modify the point of view and the point in time) to move over the scene and to move forward and backward through time to highlight changes and movements.

Manual vs. Automated Transcription

The first step in any processing of analog historical text is to convert the text into a digital format suitable for use within a computer-based system. A significant, and as yet unresolved, obstacle to working with historical texts involves the transcription process from analog text to digital text. The lack of a viable technological solution makes transcription a slow and tedious manual process with the attendant potential for the introduction of errors. In some cases, Optical Character Recognition (OCR), where a combination of computer hardware and software is used to read the texts, can be applied to automate the process. A representative software package designed to OCR handwritten text developed at the University of Innsbruck is named *Transkribus* (Digitisation and Digital Preservation group, 2018). This system works by allowing users to initially upload images of documents along with corresponding manual transcriptions. The images in combination with the transcriptions are used to "train" the system to recognize writing patterns and match them to words in a specified language. As patterns are encountered that cannot be matched, they are identified so that a person can manually enter the word, thereby "teaching" the system what word the particular pattern corresponds to. Additionally, language specific syntax and semantic rules are used to aid in determining possible words within a given context. For example, in the English language sentence "I went to the store", if the word "to" was illegible for some reason, then the system could narrow the options to a few possibilities such as "to", "on" "behind", or "in" – essentially words indicating the relative location or relative motion. The use of semantic rules, however, does not help in the interpretation of the names of people, places, or events. Because names are of particular interest to historians, this is a significant deficiency (Figure 2.1). Actions without identifiable subjects or objects may be useful for the statistical analysis of the number of occurrences of a given type of activity, but without knowing the participants it is impossible to develop an understanding of social or familial relationships.
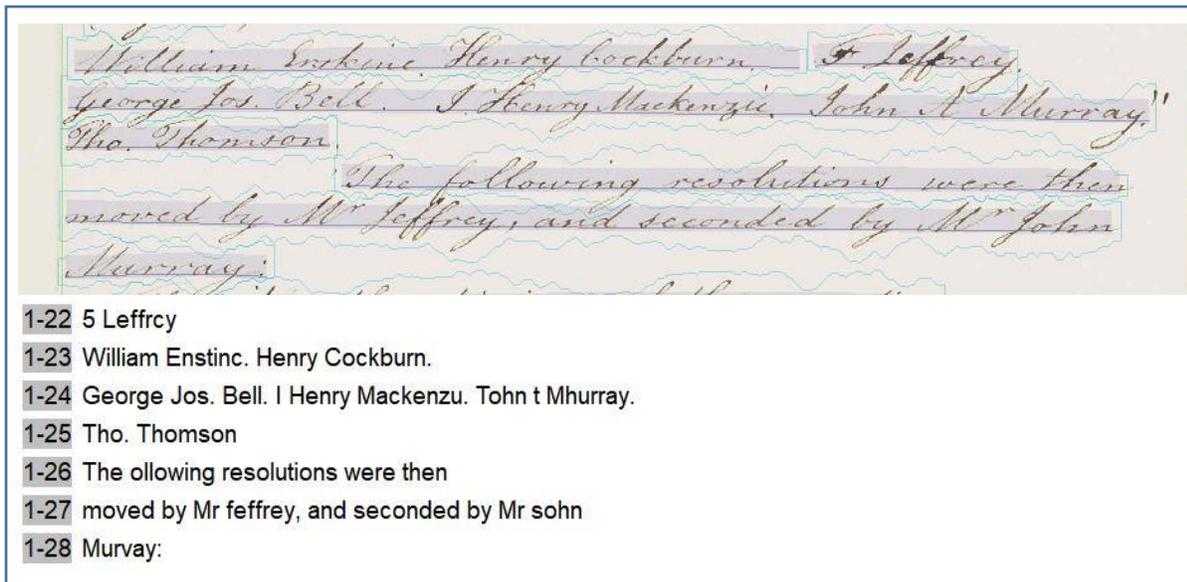
Figure 2.1: Sample OCR transcription result using Transkribus on an english language document. The upper section of the figure contains an image of the text with individual ngrams (sequences of terms) enclosed in blue lines. The lower section of the figure contains the transcribed content. Out of ten names in the original text, two were correctly transcribed.

Once the data are in a digital format the content must be reviewed to identify the significant content. For small volumes of research documents, manual techniques where the texts are read and names are recorded, may be sufficient and generally highly accurate. In this approach, the researcher uses subject matter knowledge to identify individuals and places to account for ambiguity. For large corpuses of texts, time constraints make manual content extraction untenable and automated techniques must be applied. In these cases, computational linguistics techniques, Natural Language Process (NLP), Named Entity Recognition (NER), and Corpus Linguistics are used to find, disambiguate and assign levels of confidence to the extracted place names and person names (Gregory, Donaldson, Murrieta-Flores, & Rayson, 2015; Martins, Manguinhas, & Borbinha, Extracting and exploring the geo-temporal semantics of textual resources, 2008). These techniques can be as simple as scanning the texts and identifying names that match existing gazetteers or name lists. When a matching gazetteer entry is found for a placename the

corresponding location can be assigned. For person names, the matching person in the name authority can also be recorded.

However there are several confounding issues that can arise (Gregory, Donaldson, Murrieta-Flores, & Rayson, 2015; Martins, Manguinhas, & Borbinha, Extracting and exploring the geo-temporal semantics of textual resources, 2008; Southall H. , 2003). First, multiple physical locations with the same name can lead to ambiguous place names. Second, spellings can vary for the same place; for example, the spellings, St Vincent, St. Vincent, Saint Vincent, could all refer to the same location. Third the words used for a name could mean several different things; for example, *Yorkshire* could be a place in England, a person's name, or a type of pudding. To help mitigate these issues, more advanced corpus linguistic techniques that include the incorporation of lexis, syntax, semantics, pragmatics, and discourse, must be employed (Gregory, Donaldson, Murrieta-Flores, & Rayson, 2015). These techniques can be enhanced based on additional logic such as that applied by Martins and others (Martins, Manguinhas, & Borbinha, Extracting and exploring the geo-temporal semantics of textual resources, 2008). In his experiments with text processing, he and his associates developed a system that performs additional evaluations including the use of *one referent per discourse* which for a geographic reference often means only one of its senses when used multiple times within one discourse context. Also, what is referred to as *related referents per discourse* where geographic references appearing in the same discourse context often indicate nearby locations. Finally, d*efault senses* can be assigned to ambiguous references since important places are more likely to be referenced. For example, the name *Lisbon* in a text is more likely to be a reference to a city than a street.

Deep Mapping

The spatial humanities with its focus on the importance of space and place in understanding social processes lies at the intersection of critical geography and the spatial turn – "the engagements between GISc and the humanities" (Bodenhamer, Corrigan, & Harris, 2010; Harris T. M., 2017). As Harris points out "The importance of the Spatial Turn in the humanities is to not only build on the methods that are central to geography but to also shift focus toward geographical concepts and spatial thinking in order to gain spatial insights not previously seen" (Harris, Bergeron, & Rouse, Place, spatial storytelling, and immersive visualization in the humanities, 2011). The concept of *deep mapping*, which mirrors somewhat the *thick description* of anthropology (Knowles A. K., 2014), is now coming into its own as a means of expressing these more critical ideas within a computer-based context. Pearson and Shanks (Pearson & Shanks, 2001, p. 64) suggest that,

> [deep maps] record and represent the grain and patina of a location...[through the]... juxtapositions and interpenetrations of the historical and the contemporary, the political and the poetic, the factual and the fictional, the discursive and the sensual; the conflation of oral testimony, anthology, memoir, biography, natural history and everything you might ever want to say about a place.

Deep mapping then, seeks to provide deeper connections to the places located on a map and enable the ways in which place affects people and their interaction with places to be explored. One goal of deep mapping is to delve into the consciousness of historical actors, applying concepts of positionality and reflexivity on the part of the researcher along with a consideration of how historical actors are representation and incorporated into the research (Guelke, 1997). Springett, argues that,

> Deep mapping distinguishes itself from flat ontology by attempting to address such hierarchies, which are specifically social and ecological in nature in a way that is inherently political. It takes a performative approach of undoing or decolonizing, and seeks to affect

a specific *enaction*, a process set into practice by altering or deepening peoples' perspective of place and their relationship to it—to one that is socially-just, ecologically-aware, more democratically-located, and one which erases disciplinary boundaries to extend beyond the academy. (Springett, 2015, p. 634)

Yet, successfully transitioning from *deep mapping* as a concept to practical implementation is itself an unresolved task, for as Harris (2017) contends, "Moving from deep map concept to method is not without its challenges.". The concept of deep mapping itself remains an open question for some as noted by Wood (2016, p. 15) when he states "Deep maps, deep mapping… Yes, yes, but such a strange name for the practice. A practice that so often delivers far, far less than it promises. Especially maps. So many instances of deep mapping lack any at all.". Rather than rely on an attempt at definition, it is instructive at this point to give an example to help clarify the evolution from an HGIS representation to a Deep Map. The following scenario is also offered as an example of how spatial analysis might be used to generate new lines of investigation, and new conclusions and insights.

The distinction between a traditional map and a deep map may be drawn by comparing the journeys taken to Salt Lake City by the Donner Party in 1846 with that of the Later Day Saints in 1847 (McGlashan, 1947; Stegner, 1992). The Donner Party embarked from Independence, Missouri and the Later Day Saints from Nauvoo, Illinois with the two paths intersecting at the Platte River near present day Kearney Nebraska. From Kearney, their routes would be essentially identical; both first travelled the established route to Fort Bridger in what is now northeast Utah.

At Fort Bridger, in 1846, the Donner Party were told of a new route that would cut several hundred miles off their trip. The idea was proposed by a land speculator named Lansford Hastings and has become known as the Hastings Cutoff. The uncleared route led south from Fort Bridger to the top of a canyon, now called Emigration Canyon, that opens into the Salt Lake Valley. The

party spent weeks cutting a trail through the brush and woods, turning the shortcut into a delay that would set the stage for their late season arrival and infamous hardships in the Sierra Nevada Mountains. The following year the Later Day Saints, led by Brigham Young, also took the now cleared Hastings Cutoff to Salt Lake City where he proclaimed, "This is the place", initiating the development of the region into what he hoped would be a new nation founded on Later Day Saint doctrine.

As mentioned, looking at the maps alone the route and therefore the journeys appear identical. By adding a timeline or time labels, the difference begins to come out and questions about the travel times spent in the Hastings Cutoff may come up. Yet with a map showing the route and a timeline we have only a traditional map that gives little indication of the experiences of the travelers. A possible technique that could be used to 'deepen' the map would be to analyze the diaries and memoirs of the travelers, extracting words linked to the emotions of the writers at different points along their paths. Historical documents and histories of the trips could also be used to derive a sense of the travelers' state of mind. For example, Virginia Reed, of the Donner party commented that during the first part of the trip she was "perfectly happy". There would likely be many similar indications of excitement, apprehension, anticipation, hope, possibly freedom in the case of both groups; in general, a positive feeling at the start of their journeys. At Fort Bridger, this would change for the Later Day Saints, positive words of anticipation may begin to appear more frequently as the goal began to come within reach. For the Donner Party, the first indications of dissent appear as they debate whether to stay on the established but longer route, or to strike off on the promise of a shorter therefore faster trip. Upon arrival in Salt Lake City, the Later Day Saints would celebrate, and a positive sense would prevail. For the Donner Party, the site of the

Salt Lake Desert ahead and a warning letter left by Hastings telling them that the route ahead was all but impassable would lead to a sense of despair and desperation.

With this additional information taken from historical sources a more evocative, deeper, map can be developed. A first step could be to color code the routes on the map to indicate the emotional state of the travelers ranging from positive to negative. The result would be two distinct maps that begin similarly but end in Salt Lake City with very different emotion-based appearances. For the Later Day Saints, the route would flow from initially a positive outlook to despair upon arrival.

The question to be asked at this point in the map making process is "So what?". The new maps would provide a quick and somewhat intuitive visual indication of more than just a route of travel, which is useful, but they would still be superficial in their content and not lead the viewer to internalize what was going on. In addition to color coding, a deep map would have enhancements such as interactive options that would link specific locations on the map to transcripts or images of the documents that refer to them. To develop a better sense of the journey, an author may also provide links to still imagery or video of the referenced locations as well as other landmarks that they travelers would have encountered.

In addition, one of the members of the Donner Party (sometimes called the Donner-Reed Party) was James Reed. On the trip, he brought along a large wagon holding his possessions including even large pieces of furniture. A potentially compelling deep map feature would be the additional of a link to an immersive an interactive Virtual Reality (VR) model of an actual wagon of the type that Reed used. This would provide a sense of the scale and mass of what the travelers pulled along the dirt road. The model could be set in a wooded mountain scene based on actual terrain from the region. Oxen could be included in the model as well adding not only to the sense

of scale but also an awareness of the additional effort required to manage, guide, and maintain these large animals in a wilderness environment. This visualization combined with imagery of the Wasatch Mountains would not only give a sense of the ordeal then endured but may also lead to "what-if" questions: what if they would have sold their possessions and travelled light? Would they have made the journey safely? The weeks spent clearing a road on the Hastings Cutoff solely for the purpose of allowing the wagons to make if would have been avoided. This may also lead the viewer to question more philosophically whether possessions are a gift or a burden – do we own them or do they control our actions?

A deep map would also provide links to sources describing the backgrounds and motivations of the parties. The individual motivations of the Donner Party members and the collective motivation of the Later Day Saints. Were the Later Day Saints pioneers fleeing religious persecution and being led by Latter Day Saints or were they duped being led by salesmen with a history of petty scams? A well-constructed Deep Map would provide references to arguments on both sides, allowing the viewer to develop informed opinions. With these additions, there exists a deep map that is a launching point for further research. It is a map that provides information as well as intellectual stimulation. By taking the additional step of allowing other researchers – or the general public – to add new links to documentation and imagery the map can become a living, every increasing, knowledge base.

Deep maps are, in essence, location based launching points for a broader exploration of the subject matter. To accomplish this, deep maps rely on the application of multiple (often web-based) technologies to develop, personal, and positional representations. They can display images related to the subject that may be stored locally with the map or they may provide hyper-links to

web imagery and documents and they can provide links to web pages developed by other researcher that contain additional information or more topic specific information.

In summary, the deep map does more that show lines and points on a map. It provides mechanisms that allow viewers to interact, explore, and question events. It provides deeper connections to the places located on the map and ways in which those places affected people and people altered the places. After experiencing a deep map of these two journeys, *Hastings Cutoff* is no longer just a trail and, significantly, it has very different meanings from the alternative perspectives of the Donner Party and the Later Day Saints.

Virtual Reality

A new generation of information exploration and presentation technology is embodied in Virtual Reality. In his book 'Our Mathematical Universe', Max Tegmark proposes that the physical universe is, in reality, a mathematical structure (Tegmark, 2014). Everything from the house we live in, to the planet we live on, to the sun we revolve around, are manifestations of a set of fundamental equations. Tegmark's contention is not that mathematics can be used to model the physical world, but rather that mathematics is so perfectly adept at describing the physical universe because they are one-in-the-same. While not all mathematicians subscribe to this idea, the notion that the physical world can be described and, importantly, modelled using the language of mathematics is broadly accepted (Frenkel, 2014).

The ability to model the physical world mathematically is the foundation on which Virtual Reality (VR) 3D immersive technology are built. Shapes of 3D objects can be modelled as a set of polygons and using simple geometry, it is possible to make those polygons translate and rotate in space. Through scaling and perspective calculations, the objects can be made to appear to move

towards or away from a given viewpoint. In immersive development, equations representing Newton's laws of physics can be applied to simulate the interactions of objects when they encounter one another. Virtual objects can be made to rebound from collisions or shatter depending on how the structures of the simulated materials are modelled. Because we are accustomed to a world that can be described by these same equations, we perceive the actions and interactions of the simulated objects as 'realty'.

While much of the emphasis of virtual reality is towards entertainment, commonly gaming, the technology lends itself well to more serious applications. Immersive virtual reality is inherently spatial; the experience offers the ability to look around a local environment or to transport across the galaxy. The combination of movement and sensory stimulus is a natural fit for the field of geography. Physical geographers and GIScience specialists can examine spaces and analysis the potential interactions of objects and individuals. They can develop a sense for the movement of objects and people as the encounter physical obstacles or pathways. While not seeking a perfect reflection of reality, humanists can examine how a physical space can take on a cultural or personal meaning by immersing themselves in the sights, sounds, and places that others have experienced in life. Humanists can develop a sense for the challenges of a person with a physical handicap trying to navigate a busy city or the sense of identity that a native community can take from the surrounding landscape.

By utilizing immersive Virtual Reality, historical research can be extended further to offer a greater visualization of events as we know them now along with a better sense of place at the time that the events were unfolding. For the fields of history and humanistic research writ-large, immersive technology can be a tool for research, education, and preservation. As discussed above,

researchers could benefit through interactive collaboration and the development of visualizations in order to gain greater insight into the context of events in the past. By incorporating the stimulation of other senses by through temperature, scent, sound, touch, or lighting, very compelling scenes can be developed (Harris T. M., 2017). Cognitive studies have demonstrated that the ability to *feel* and experience an event rather than simply memorize it can lead to a greater ability to retain the information and a greater ability to sense patterns in actions across time and space (Marshall, 2007).

In addition, historic sites and areas of archaeological significance are often impacted as human population continues to grow and land is developed for housing or commercial purposes. Given this situation, the capture of detailed 3D representations of threatened sites becomes increasingly important as a means of preserving the past (Lipe, 2000; Nickens, 2000). Once sites are gone, virtual reality may be the only means by which we can view what previously existed. Even where the site itself has been preserved, as in the case of many U.S. Civil War battlefields, the surrounding landscape may be dramatically altered. This loss can significantly detract from our ability to interpret the historical context of events. By simulating the contemporary landscape in an immersive setting, historians may be able to capture a better sense of how and why events of the past unfolded as they did.

In a previous section, the technique of applying HGIS to supplement historical narrative was introduced using the examples of Lee at the Battle of Gettysburg and the prehistoric hillfort at Danebury. While these applications of HGIS technology enable us to gain insight into some of the rationale behind events, gaining a deeper sense of place requires a more realistic and interactive experience on the part of the observer or researcher. With the advent of immersive 3D capability, this type of research can be extended to offer a greater visualization of events as we know them

now and also a better sense of place at the time that the decisions were made. For example, rather than imagining what Lee could see at Gettysburg, immersive 3D offers us the ability to place ourselves virtually in the Seminary, to look through the trees, and to squint into the sun. Rather than looking at visible areas and blind spots as shades of grey and brown on a viewshed map it can allow us to 'see' the terrain from the perspective of Lee, giving us an even greater insight into his decision. In the case of the Danebury hillfort research, not only could we see a map of the time-to-travel to reach a safe location, we could place ourselves virtually in a farmstead and gain a sense of the effort required. We may also be able to gain additional insight into how the inhabitants perceived their situation; how they envisioned alternate routes of escape based on an enemy's possible arrival paths, on obstacles and on nearby hiding places. We may even be able to gain some sense of the level of vigilance that they had to maintain as they went about their daily lives.

Immersive 3D technology supports these types of large-scale narratives through viewer centered visualizations and interactions either from specific ground points or from bird's eye perspective. In addition, it also opens a world of intimate geographic visualization of history not available from traditional cartography. Immersive technology can be applied to virtual reconstructions of smaller areas and features such as temples or villages or rock shelters, and to place the viewer within them (Wernke, Adams, & Hooten, 2014). Environmental conditions such as ambient lighting, weather, and vegetation can be altered to provide the viewer with varying perceptions of place.

VR Technology

The data requirements for presenting a high-resolution 3D image, even of a small are, are considerable. These data must be collected, cleaned, stored, retrieved and manipulated by means

that are cost effective and timely. Fortunately, technology has advanced to the point that this is possible even with the use of off-the-shelf components (ESRI, 2019). Gathering point cloud data using lidar and stereo photography has become mainstream. Aerial data can be gathered using fixed wing aircraft, helicopter, drones, and terrestrial systems. While in many cases this is still preferable because of the ability to control collection parameters, there is an ever growing base of pre-collected data that is available from a variety of governmental and private organizations such as Assessorato Territoria e Ambiente Valle d'Aosta (Assessorato Territoria e Ambiente Valle d'Aosta, 2014) and the U.S. Geological Survey (U.S. Geological Survey, 2014).

Similarly, 3D immersive hardware has also become more ubiquitous and is available at a variety of scales. The Computer Automatic Virtual Environment or CAVE, is room sized viewing area with three to six viewing surfaces (Oculus, 2019; Visbox, 2019). This type of environment is suited for both individual researchers or for several collaborative researchers. Multiple viewers can experience a visualization at the same time while discussing points of significance. At the other end of the 3D hardware spectrum are the head-mounted display systems (HMD) (HTC Inc., 2019). These devices offer relative ease of use and significant cost savings, with systems ranging in price from several hundred dollars to several thousand. With a head-mounted system, individual researchers can examine and manipulate imagery and researchers can also collaborate either locally or remotely.

Software to support immersive 3D systems include data analysis, image rendering, simulation, editing, environmental simulation, and the presentation of supplementary information in the form of augmented reality. Currently available software allows users to manipulate and edit 3D imagery in real-time and supports the creation of simulated movement of figures through a scene and simulated changes in environmental conditions such as weather or lighting (Unity

Technologies, 2018). Along with this, VR applications can be stored in on-line repositories for access by scholars and researchers (Valve Corporation, 2018).

Digital Data Curation

Data collected and produced in the course of humanities projects are often considered by-products and essentially abandoned upon project completion and publication. However, rather than being by-products these data are artifacts with value in-and-of themselves (Heidorn, 2008). Importantly, the intrinsic value of the data extends beyond the scope of the original research project and beyond the end date. As with other physical artifacts maintaining the long-term viability and accessibility of digital data is now viewed as an important aspect of historical research planning. However, awareness of the importance of curation planning and how best to curate, archive, and preserve digital data has not filtered down to all research projects (Rothenberg, 1999).

To support the advancement of digital data curation, the roles of research libraries and archives is changing within the scope of organizational research projects as they seek to position themselves as active agents in the research process and as consultants and resources for curation planning (Walters & Skinner, 2011). This support covers all aspects of the research process from raw data collection, to data cleaning and process, to data analysis, to publication (Kong, 2016). The usefulness of this service is apparent in the recent requirement for specific data management plans as part of NEH and other grant proposal requirements. Examples of this type of institutional support are given by Walters (2011) with specific projects such as the Valley of the Shadow and general support for the development of standards like TEI. However, in a survey of Australian research institutions to determine the level of awareness of researchers, it was shown that 82% of project teams involved in any type of research simply planned on storing their own data as their

means of long-term preservation (Kennan & Markauskaite, 2015). This laxity may indicate either an un-willingness to use organizational resources or, more likely, a lack of awareness of the importance of data curation. Compounding the problem is that even among the professions closely involved in preservation there is a broad lack of understanding of curation standards and processes. The Preservation Metadata Interchange Standard (PREMIS) specification itself states: "Most of the staff in libraries, archives, museums and other cultural heritage organizations don't have any direct involvement in digital preservation. In that case, it's enough that you know what PREMIS is: a data dictionary for preservation metadata" (Caplan, 2009, p. 6).

In discussing digital data curation, a distinction must first be made between digital and analog data. Digital data for purposes of this discussion is a recording of 1's and 0's on some type of computer-based media. Analog data is a physical object such as a book or a painting or a stone tool. Analog can also refer to the format used to record information, often audio, on magnetic tape. In another interpretation, the term analog can be used to describe structured digital content that has been converted to an image, as in the case where a screen capture is taken of a Word document – the Word document would be called the digital version and the screenshot the analog version. However, for current purposes digital is defined as anything stored on computer-based media and analog is defined as a physical artifact.

Another key concept in understanding digital data curation is that of metadata, or data that describes data. A familiar type of metadata that is not often recognized are the simple filename extensions used commonly on personal computers. For example, when a file has a .pdf extension, the computer operating system knows to read the file using Adobe Reader or when it has a .docx extension it knows to use MS Word to read the file. If the file extension is removed or changed, the bits contained in the file become effectively unusable unless other steps are taken. Along with

defining how data are interpreted, metadata may also define other aspects such as ownership, source, provenance, preservation requirements.

The basic objective of any curation activity is to preserve the long-term viability and accessibility of an item of current or perceived interest. In the case of digital data, achieving this objective is a multifaceted process with many attendant issues. The first issue is that with digital data there is nothing tangible to examine therefore all access to the data must be mediated by some type of software rendering process. The 1's and 0's that make up digital data have no inherent meaning on their own. They must be read by software to be interpreted in a way that gives them meaning and so that they can be rendered for use by humans. The additional implication is that the curation of digital data must be accompanied by the curation of the rendering software as well.

A related issue is that digital data can, over time, become inaccessible, or in the words of Rothenberg, *"Old bit streams never die—they just become unreadable"* (1999, p. 2). There are three primary factors that can cause digital data to become unreadable; media degradation, software obsolescence, and hardware obsolescence. Media degradation simple means that the disk or tape that data are recorded on deteriorates to the point where the 1's and 0's can no longer be read consistently. This can lead to a complete loss of access to digital files or the corruption of portions of data contained within files. Software obsolescence means that the software (or the specific version of a software package) that was used to create the data is no longer available. An example of this is when one tries to open an old spreadsheet only to find that it cannot be read by the current installed version of software. Hardware obsolescence occurs when the physical computer or reader required to access the physical media on which the data are stored are no longer available. A notorious example of this occurred with the 1960s decennial census data. These data were originally recorded on UNIVAC Type II-A tapes, however by the 1970s these tapes were

obsolete and only two drives were in existence in the world. The data were copied to other media but not without significant effort and expense. The Committee of Records of Government later stated that "the United States is in danger of losing its memory." (RLG - OCLC, 2002).

Another factor that affects digital data accessibility is that of intermittent data loss. When archiving analog objects, the best practice can often be benign neglect; as long as environmental conditions are appropriately maintained it might be best to leave the artifacts alone so that physically handling them does not cause them to degrade. For digital data this is not the case for the curation of digital data must be viewed as a process (Abbott, 2008). Degradation and obsolescence occur over time and data must the constantly monitored for quality and fixity. Corruption of data can be introduced as part of a standard back process when errors occur during reading or writing; a condition is sometimes referred to as 'bit rot'. Because of this, checks must be performed with each backup to ensure that the original and the new copy match exactly. When discrepancies are found it must be determined which version is correct. Procedures must also be enacted to ensure that the data are kept in formats that are accessible using current computer technology. In addition, another confounding issue regarding digital data curation is that there is no inherent authoritative version of digital data. As soon as digital data are duplicated or are 'loaned out', multiple copies exist, each of which is indistinguishable from the original. When this occurs, there is no intrinsic way to know which is the original. The authority of a version of digital data is therefore based on the authority of the curating institution not the data itself. To help mitigate this issue the concept of the *Trusted Digital Authority* (RLG - OCLC, 2002) has been developed. Trusted digital authorities are repositories that are certified as the source of reliably authentic versions of given data sets.

Impediments to the interchange of digital data due to its ephemeral nature also exist. For digital data to be of use it requires metadata for correct interpretation and, in addition, it requires metadata in order to be discovered and exchanged. Because of this, standards have been developed that define all aspects of related to the management of digital data. These include metadata that describe the content of the data, metadata that describes how to read and render the data, metadata that support data discovery in clearinghouse environments, metadata that facilitate transmission over the internet, metadata that support data administration, and metadata that support curation (Caplan, 2009; NISO, 2004; OCLC/RLG Working Group on Preservation Metadata, 2002). An understanding of the role and implementation of each type of metadata is therefore critical to the successful exchange of digital data.

Along with general issues related to the curation of digital data, operational activities must also be considered. Significant work has gone into the development of procedural standards for operating a digital archive and the creation of metadata standards that support archival information management (CCSDS, 2012; OCLC/RLG Working Group on Preservation Metadata, 2002; RLG - OCLC, 2002; Caplan, 2009). These standards are comprehensive and are designed to be understood by archivists or commercial entities providing archival services. These standards, however, describe what needs to be recorded and what operations need to take place but not on how to implement them. Implementation is the difficult and expensive part of establishing an archive. To support this there have been several open source archival packages, such as Fedora, that implement these standards along with third party commercial service providers (Moore, Rajasekar, & Marciano, 2007; Walters & Skinner, 2011).

Curating Geospatial Data

Geospatial data are a special class of data that have unique characteristics that distinguish them from other digital data and therefore require additional consideration for their curation (Bleakly D. R., 2002). Geospatial data require additional metadata to define coordinate systems, projections, and datum and they may also have to record combinations of raster and vector data at varying levels of granularity or resolution. Spatial data also often exist in proprietary formats that do not conform to any formal interchange standard; with ESRI shapefile having become the de facto format. Unlike text or tabular data, geographical data are generally represented as maps which have no inherent formatting. Issues of positional accuracy must also be addressed such as those introduced by errors induced by manual measurements, skill level of the technicians acquiring the data and those processing it, original intent for the data. Geospatial data may also define the connectivity between data elements in the form of networks of features, and spatial relationships such as adjacently, overlap, and intersections (McGarva, Morris, & Janee, 2009).

Because of these idiosyncrasies, the curation of spatial data requires specialized metadata to adequately describe them. In response, there has been a great deal of effort put into the development of standards and standards bodies to ensure consistent support for spatial data curation and access. This includes work being done by organizations such as the Open Geospatial Consortium (OGC) and the Federal Geographic Data Committee (FGDC) and metadata standards (Metadata Ad Hoc Working Group, 1998; Open Geospatial Consortium, 2017). Research and academic libraries also understand the distinctive characteristics of spatial data and are working to develop processes that support collection, analysis, preservation, publication, and discovery of these types of data (Kong, 2016). Because of the complex and often proprietary nature of digital

data in general and spatial data in particular, the development of authoritative format registries, such as PROMON the Unified Digital Formats Registry (UDFR)continues to evolve. These registries provide the critical link between the data and the software systems that are used to access them (Walters & Skinner, 2011).

Several metadata standards have been developed to specifically meet the requirements of spatial data. These include the Spatial Data Transfer Standard (SDTS) (Bleakly D. R., 2002), which was designed to provide a non-application specific format for the interchange of spatial data. This specification relies the encoders and decoders which convert between the proprietary formats and the standard. Alternatively, the Federal Geographic Data Committee (FGDC) (Metadata Ad Hoc Working Group, 1998) standard is a metadata standard that was designed to facilitate the interchange and access to various formats of spatial data. The currently accepted 'global' standard for spatial data interchange is the 2003 ISO 19115 format which drew on the earlier FGDC standard (McGarva, Morris, & Janee, 2009). Current metadata standards, such as the Metadata Encoding and Transmission Standard (METS) for data interchange and Preservation Metadata Implementation Strategies (PREMIS) for curation, and the ISO 19115 standard for spatial data are adequate in all but a couple of areas. They are adequate in that they define far more metadata attributes that would normally be required by any given project for data curation and interchange. METS, for example, supports the embedding of other metadata within it as well as domain specific metadata extensions, effectively broadening the applicability of the standard. However, within the current spatial metadata standards there is no standard for identifying connectivity between data elements, networks between features, and spatial concepts like adjacently, overlap, intersections of data. Since many GIS processes involve spatial analysis that

examine these types of relationships this information must be maintained in annotations or dataset wide descriptive tags.

CHAPTER 3 - Building the Spatial Historian System

The Spatial Historian does not presume to be a replacement for historical research or GIS. Just as a word processor cannot make a writer nor a spreadsheet make an accountant, this system will not turn poor research into good history. This system, left to its own devices, will produce exactly nothing. The Spatial Historian can, however, help the historian capture and organize large volumes of data, both internal to a project as well as external, and thereby provide a comprehensive view of a research project that can also be decomposed to the smallest level of detail. Just as a spreadsheet platform can enhance the tabular data through the visualization and production of charts and graphs, so too the system described here offers a unique method of visualizing historical events as they occur over space and time. Where the user of a spreadsheet can slice, filter, and compare tabular data, so the Spatial Historian allows historical events to be similarly filtered by time and geography and to be viewed in conjunction with one another. Essentially, the role of the Spatial Historian is to convert historical source materials into historical information.

Developing IT infrastructure for STEM or for humanist needs differ little from a hardware perspective. While the processing speeds and storage capacity may very between hardware systems, digital computers in general perform the same processes. There is a difference, however, between the software available to STEM and humanist needs. Software for the humanities must be critically informed and must offer a "visual narrative", supporting the literary narrative of traditional humanities (Burdick, Drucker, Lunenfeld, Presner, & Schnapp, 2012). Furthermore, for historians who do not have a technical background, a research system cannot be a disjoint

application of disparate off-the-shelf software packages requiring a knowledge of how to transition data from one to another and how to combine the specialized output from each into a cohesive presentation. In the words of Martyn Jessop (2004, p. 348) "Where the use of a GIS is required, the unavoidably steep learning curve can be minimized by utilizing the interoperability of GIS software with existing software that one is familiar with, for example by keeping data in Microsoft Access or Excel and linking it to the GIS." Yet embedded within this statement is the expectation of a knowledge of the individual software and how they might be linked to a GIS. For historians this is not always achievable, and without a technical background, humanists may not be best positioned to develop software themselves and may even have difficulty defining exactly what a useful software system would look and act like in practice.

Historical research is an exploratory process and the development of a system platform like the research itself, is an evolving process. A fluid approach to software development is contrary to the rigid and pre-planned requirements process that is fundamental to most mainstream software development. The nature of software development should ideally be reevaluated in response to the nature and needs of humanistic research. The platform development process should be an agile and interactive process whereby technical design is tested and refined against the language of humanistic research (Burdick, Drucker, Lunenfeld, Presner, & Schnapp, 2012). As such, symbolic representations and interactive experiences should evolve through open and ongoing feedback between the software developer and the researcher. The development of a system such as the Spatial Historian must walk a path between mediating the technology and the historian and thereby translate the humanistic needs of the historian into code and a computer system that satisfies those needs.

To successfully support historical research and fulfil the promise of the digital humanities, the Spatial Historian provides a number of computational functions. Among those functions are data capture capabilities such as document scanning, digital audio recording, and digital image recording. Once captured, the system provides data management capabilities in the form of file services and database systems. The extraction of historical content related to people, places, and events often requires text processing capabilities including optical character recognition (OCR) to convert document images to digital format, and digital heuristic text processing techniques along with the application of gazetteers to parse that information. The system should also support the construction, storage, and presentation of alternate media formats including video, audio, maps, network diagrams, and even 3D virtual representations.

The Spatial Historian System Architecture

The Spatial Historian implements a multi-tier architecture which is composed of three layers of logical processing and is comprised of a data tier, an application tier, and a presentation tier (Figure 3.1). This architecture provides many benefits for the production and development environments by compartmentalizing the user interface, business logic, and data storage functionality. Doing so allows specific parts of the system to be updated independent of the other parts and this reduces development cycle times by offering the ability to replace or upgrade independent tiers without affecting the remainder of the system. For example, the user interface can be updated based on user feedback without affecting the underlying functional business and data access logic. This architectural approach is also useful for integrating third party software, such as the Microsoft Azure Speech-to-Text (Microsoft, 2018) service, into the application.
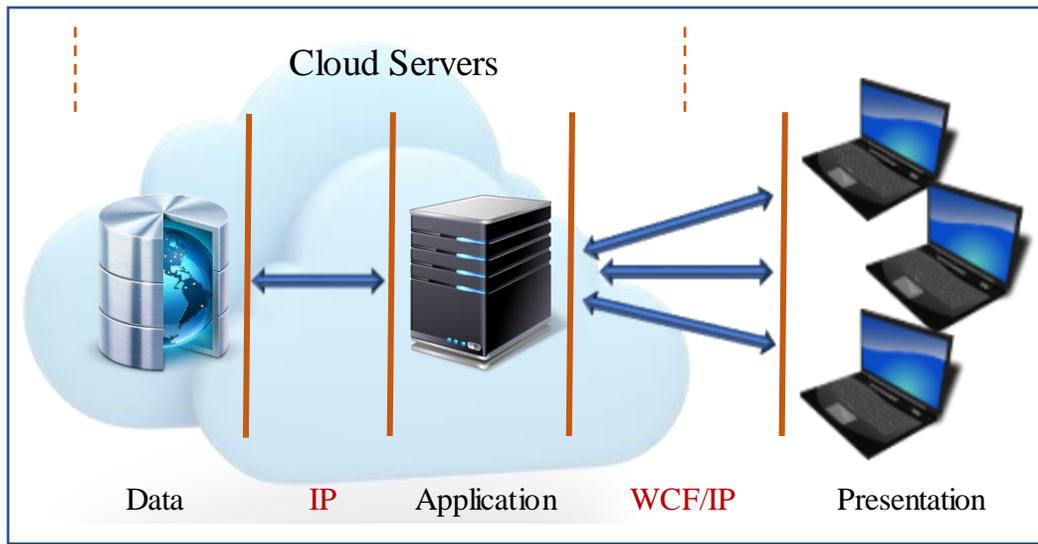
Figure 3.1: Three tier system architecture with the database and application services running on cloud servers. SQL Server IP communication between the database and the application services, and WCF over IP between the application services and end-users.

Another advantage of this architecture is that it is scalable and each layer is capable of being scaled independently depending on the processing demands at any given time. For example, as larger quantities of documents are transcribed, the data tier can be upgraded to support the additional storage and performance requirements. Additionally, the independence of the tiers offers flexible deployment. During the initial development of the Spatial Historian, for example, the data and application services were hosted on a server housed in the West Virginia GIS Technical Center at West Virginia University. However, as the system was adopted for production work outside of the university, the services were moved to Microsoft Azure cloud servers (Micorsoft, 2018). Because of the multi-tiered architecture, this migration did not require any changes to the software. The three-tier architecture includes a data tier, application tier, and an application tier. The three-tier architecture allows the Spatial Historian to run in either Production mode or Test mode without requiring any programming changes. At login, the user is given the option to choose the desired mode and then, based on the selection, the presentation tier sends the

mode selection to the application tier as part of each request. Using this information, the application tier dynamically determines which data tier (database) to access.

The data tier houses databases where data and media are stored and managed. Operations in this tier include database updates and retrieval, along with database security administration and data backups. The data in this tier are only accessible through the application tier and cannot be directly accessed by the end user. The databases themselves are hosted on Microsoft Azure SQL Servers which support spatial data types used to store geographical information and standard data types used to store attachments, transcription related details, and ancillary data.

The application tier, or middle tier, controls system functionality and performs data manipulation logic. This tier acts as both a firewall and the conduit between the user's interaction and the data stored in the data tier. This tier also performs user authentication and activity logging.

The presentation tier, commonly referred to as the user interface (UI), occupies the top level tier and displays information derived from the application tier. Communication between the application tier and the presentation tier uses Microsoft Windows Communication Foundation (WCF) technology over standard Internet Protocol (IP) connections. WCF handles all the serialization and de-serialization required to transport data and function requests across the internet. This mechanism is transparent to the software developer thereby making calls to requests to the application tier appear as any other programming function call. The application and presentation tiers are designed to make the complexities of the underlying data model transparent to the user.

Software Development

The Spatial Historian is developed using the Microsoft .NET Framework. The .NET Framework is Microsoft's development platform for building applications for the web, Windows, Windows Server, and Microsoft Azure. The .NET Framework consists of the common language runtime (CLR) and the .NET Framework class library, and implements a broad range of functionality along with support for many industry standards such as XML document manipulation, HTTPS (web) communications, and database interaction. This flexibility supports the development tasks required for the Spatial Historian.

The primary coding language used in the development of the Spatial Historian is Microsoft's C#. C# is a general-purpose, object-oriented programming language that supports software engineering principles such as strong type checking, array bounds checking, detection of attempts to use uninitialized variables, and automatic garbage collection. C# exploits the functionality provided by .NET, and its robustness and portability make it well suited for developing the user interface and processing logic for both the application and presentation tiers.

```
/*-------------------------------------------------------------------------------------------
 * Method used to determine the persons sex based on Given Name...
 * -----------------------------------------------------------------------------------------*/
public static ListGender DetermineGender(string name)
{
    string[] nameParts      = name.Split();
    NameLookup nameLookup    = Globals.nameLookup.Where(l=>l.NameTypeID == (int)Enumerators.NameType.GivenName
                                            && string.Equals(TextHelper.RemoveDiacritics(l.Name),
                                                    TextHelper.RemoveDiacritics(name).Trim(),
                                                    StringComparison.InvariantCultureIgnoreCase))
                                    .FirstOrDefault();

    if (nameLookup == null && nameParts.Length > 1)
        nameLookup          = Globals.nameLookup.Where(l=>l.NameTypeID == (int)Enumerators.NameType.GivenName
                                            && string.Equals(TextHelper.RemoveDiacritics(l.Name),
                                                    TextHelper.RemoveDiacritics(nameParts[0]).Trim(),
                                                    StringComparison.InvariantCultureIgnoreCase))
                                    .FirstOrDefault();

    int genderID            = nameLookup == null ? (int)Enumerators.Gender.Unknown : nameLookup.GenderID;

    return Globals.listGenders.Where(g=>g.ID == genderID).FirstOrDefault();
}
```

Figure 3.2: An example of C# coding of a function that determines a person's sex based on the given name. This function bases its results on the name without regard to diacritics or letter case.

Table 3.1 summarizes the coding and database framework required to build the Spatial Historian as it exists at the time of this writing, the following sections provide greater detail regarding the development techniques applied at each tier. It should be noted that the coding base for the Spatial Historian is equivalent to 3,702 pages of C# code and 280 pages of XAML code at 40 lines of code per page.

Tier	C# Files	Lines C#	XAML Files	Lines XAML	Object Classes	Windows	Database Tables
Data	6	53,798	--	--	93	--	157
Application	5	10,121	--	--	--	--	--
Presentation	122	84,195	101	11,285	42	76	--
Totals	**133**	**148,114**	**101**	**11,285**	**135**	**76**	**157**

Table 3.1: Source code and database tables produced during development of the Spatial Historian

The Data Tier

The database, implemented using Microsoft Azure SQL, is the foundation for the system for all other components exist to support the entry, editing, and display components of the system. Therefore, it was critical to the success of the system that the structure of the database and the

tables it contains are sufficiently comprehensive to support all aspects of data processing across the broadest array of projects possible. The requirement of a robust database design is in stark contrast to many databases used for historical research which are often based on spreadsheet or small-scale database technologies such as Microsoft Access. However, the large-scale capabilities of this database and the data model implemented require a knowledge of SQL database administration and database design that are not a common component of a humanities researcher's toolset.

The administration of the database, even though relatively low, is performed using a browser interface provided by the Microsoft Azure platform. This interface allows the processing power and storage space allocations to be adjusted based on workload demands. The interface is also used to assign permissions that define who can access the database and what functional privileges are assigned to each user. In addition, reports are available through the interface that show system usage, data throughput, and the amount of available memory space remaining. Also, users with appropriate access privilege can connect to the database using the standard suite of Microsoft database tools including the SQL Server Management Studio which is used to design and manage the table structure.

The Spatial Historian database is made up of 157 tables, 47 of which are lookup tables created to enforce consistency during data entry. The more fundamental tables in the model, support information about projects, documents, people, places, and events. In addition, the model contains tables that track administrative information such as usernames and passwords, user permissions, the acceptance of the End User License agreement, and system errors. The 47 lookup tables are critical for maintaining the integrity of the data entered by users and are used to create

dropdown lists presented in the user interface that ensure that only valid options are recorded. For example, the selection of document content types is limited to a Text, Image, Video, Audio, Digital Image, Map, or Painting. If this list was not available and system users were required to enter the content type in a freeform text field, then the potential for typographical errors or the entry of irrelevant types is distinctly possible. Thus, a mis-typing of *Text* as *Test* is easily envisioned, causing statistical errors and potentially causing processing errors where the type of content affects how the document is processed. Because the system is designed to support multiple research projects with differing information requirements and multiple users, the content of many of the lookup tables can be configured to suit specific objectives. For example, the choice of Ethnicities available for a project related to the Atlantic slave trade might be irrelevant to a project on early-modern European history. Users with project administration permission are provided with access to user interface components that allow all project specific lookup tables to be adjusted as needed.

Because this system was developed with the ability to search for and analyze data across projects, an additional capability has been designed into the database that allows project specific variations in the lookup tables to be standardized during data export or when performing cross-project comparisons. In this manner one project may have a list of occupations with a portion of its hierarchy defined with a generic category called *Artisan* that contains the specific occupations named *cigar maker* and *weaver*. Another project may have a list of occupations with a category called *Craftsman* containing the occupations *cigar maker* and *textile worker*. When searching across projects it may be desirable to have a search for *weaver*s from the first project return a list of all *textile workers* from the second project. To support this capability the data model has three tables named *ProjectVocabulary*, *ProjectVocabularyMapping,* and *ListListTable* that facilitate this

type of dynamic interpretation of project specific content. The implementation of these tables is shown below in Figure 3.3. The structure shown in this extract from the overall data model allows the list of people's characteristics, such as skin color, to be matched across projects.
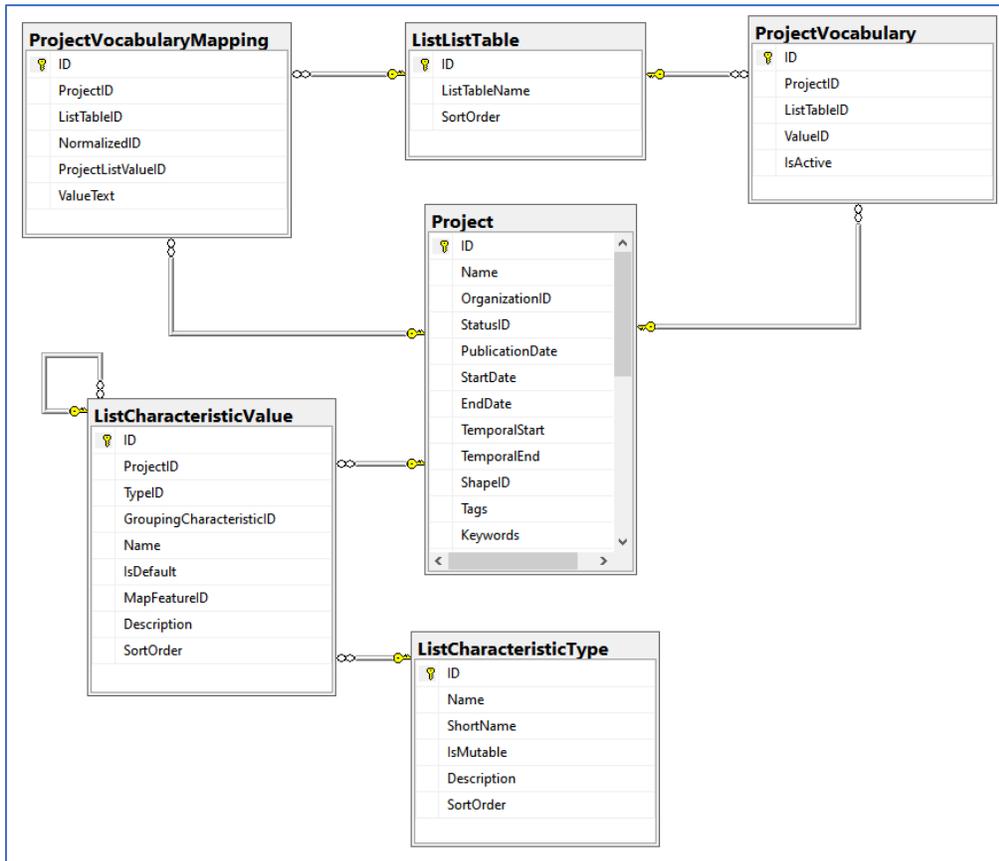

Figure 3.3: Database tables implemented to support cross-project mapping of characteristics.

The Application Tier

The application tier is also developed exclusively using C# and is hosted on a Microsoft Azure Cloud server. As with the data tier, implementing the application tier functionality on a cloud server minimizes the administrative overhead while at the same time providing high availability. As processing demand increases so the available processing power can be increased as needed. Within the context of the Spatial Historian

system, only the application tier has direct access to the database. This model insulates the presentation layer from the details and workings of the database and at the same time provides a layer of security by not requiring or allowing system users to have direct access to the database. The majority of the database requests from the application tier are performed using Language Integrated Query (LINQ) syntax. LINQ is a programming interface and technology infrastructure that supports in-line querying of a variety of list types including the result sets produced by database queries. The use of LINQ allows database access syntax within C# code to be performed as a seamless adjunct to standard C#. LINQ also provides an additional layer of database security by eliminating the potential for a class of database attacks that called *SQL Injection*. The C# code in Figure 3.4 presents an example of a LINQ statement used to query the database and to build a list of people related to a specified individual.

```
return (from relationship in db.Relationships
        join relative in db.Persons on relationship.PersonID equals relative.ID
        where relationship.PersonID == person.ID
        select relative).ToList();
```

Figure 3.4: An example of a LINQ query that returns a list of relationships for a specified person.

The application tier has also been developed to implement code that reduces the programming overhead that would normally be required to access the large number of tables in the Spatial Historian data model. Under normal circumstances, an update to a single database table would require syntax like that shown in Figure 3.5.

```
public string UpdateItemPerson (ItemPerson itemPerson)
{
    SpatialHistoryDataContext db = GetDBContext();
    try
    {
        db.GetTable(itemPerson.GetType()).Attach(itemPerson, true);
        db.SubmitChanges();
        db.Dispose();

        return "";
    }

    catch (Exception ex)
    {
        SystemLog log = new SystemLog() {
                                        SeverityID  = (int)LogSeverity.Error,
                                        Source      = "System",
                                        Username    = "System",
                                        Message     = "Processing ItemPerson Update Request: " + ex.Message
                                    };

        try
        {
            db.SystemLogs.InsertOnSubmit(log);
            db.SubmitChanges();
        }
        catch (Exception ex2)
        {
            return ex2.Message;
        }

        return ex.Message;
    }
}
```

Figure 3.5: C# function that updates the database with the information in a specified ItemPerson object.

In the example presented in Figure 3.5, the table named *ItemPerson* is being updated with the values stored in the object named *itemPerson*. If a different table, named *ProjectNote* for example, existed that required updating a similar set of code would be created with the same syntax but with different table and object class names. With the large number of tables in the Spatial Historian data model, this duplication of code becomes unwieldy and prone to errors. To clean up this situation, a helper class has been implemented in the application tier that allows updates to be performed against any database table by simply specifying the object whose data is to be updated. The helper class dynamically determines the class of object along with its corresponding database table and performs the appropriate update. The code for the helper classes generic update function is shown below in Figure 3.6.

```
public string Update(SpatialHistoryDataContext db)
{
    return Update(this, db);
}

public string Update<T> (T dataObject, SpatialHistoryDataContext db)
{
    try
    {
        db.GetTable(dataObject.GetType()).Attach(dataObject, true);
        db.SubmitChanges();
        db.Dispose();

        return "";
    }

    catch (Exception ex)
    {
        SystemLog log = new SystemLog() {
                                            SeverityID  = (int)LogSeverity.Error,
                                            Source      = "System",
                                            Username    = "System",
                                            Message     = "Update Request: " + ex.Message
                                        };

        try
        {
            db.SystemLogs.InsertOnSubmit(log);
            db.SubmitChanges();
        }
        catch (Exception ex2)
        {
            return ex2.Message;
        }

        return ex.Message;
    }
}
```

Figure 3.6: Generic function for updating the database based on the object class specified at runtime.

The helper class is only a few lines longer than a table specific update, but it is written just once. Using this generic function, an update to the ItemPerson table, or any other table, can be performed using a simple single line syntax as in the following:

```
return itemPerson.Update(GetDBContext(context));
```

As noted above, the application tier acts as an intermediary between the user interface of the presentation tier and the data access logic of the data tier. Because the presentation layer runs on the user's desktop and the application tier runs on an Azure cloud server, a communication

mechanism must be implemented that supports a seamless coding syntax along with the data handling and protocol navigation required to facilitate communication. As mentioned above, this functionality is supported using Windows Communication Foundation (WCF) technology. Configuring the interface requires specifying the IP address of the WCF service on the cloud server. When this is done, the definitions for all available service calls are made available for use within the C# development environment. Invoking the ItemPerson update function on the server from the presentation tier source code requires only a single line of code, as in the following snippet.

```
RequestResult   = Globals.serviceClient.UpdateItemPerson(Context.instance, sourceReference.GetShallowCopy<ItemPerson>());
```

The Presentation Tier - User Interface

The design of the of the WPF applications comprising the presentation tier is based on XAML, which is a declarative markup language that supports the creation of user interfaces (UI) for .NET Framework applications. XAML makes it possible to create visible UI elements in the declarative XAML markup, and to then separate the UI presentation from the application logic by using code-behind files developed in C#. XAML directly represents the instantiation of UI components defined in .NET assemblies and their linkages to related processing logic. For example, Figure 3.7 shows the window used to edit geographical coordinates for individual map features. Circled in red in the figure is a single input field that allows the user to enter the latitude. The location, format, and actions for the input field are defined by the XAML snippet shown below the screenshot. The code indicates that the input field is inherited from the base .NET *TextBox* class and its name is *txtLatitude*. The name of each item, or control, on a window is used by the

underlying C# code when implementing related functionality. The XAML also defines the location of the field on the screen as being 70 points from the left margin of the window and 74 points from the top and specifies the width of the textbox to be 78 points. In addition to these visual characteristics, the XAML code defines the actions, implemented in C#, to be performed based on user interaction. In this example, the C# function named *txtCoordinate_GotFocus* is executed when the user first enters the field. Then, for each depression of a key on the keyboard the *txtCoordiate_PreviewTextInput* function is executed and each time the user releases a key while in the field the *txtCoordiate_KeyUp* function is executed. Similar XAML code exists for every input field, piece of text, button, and image on the window as well as for the window itself. In total, this relatively simple window is defined by 70 lines of XAML code with 171 lines of C# code used to implement its functionality.
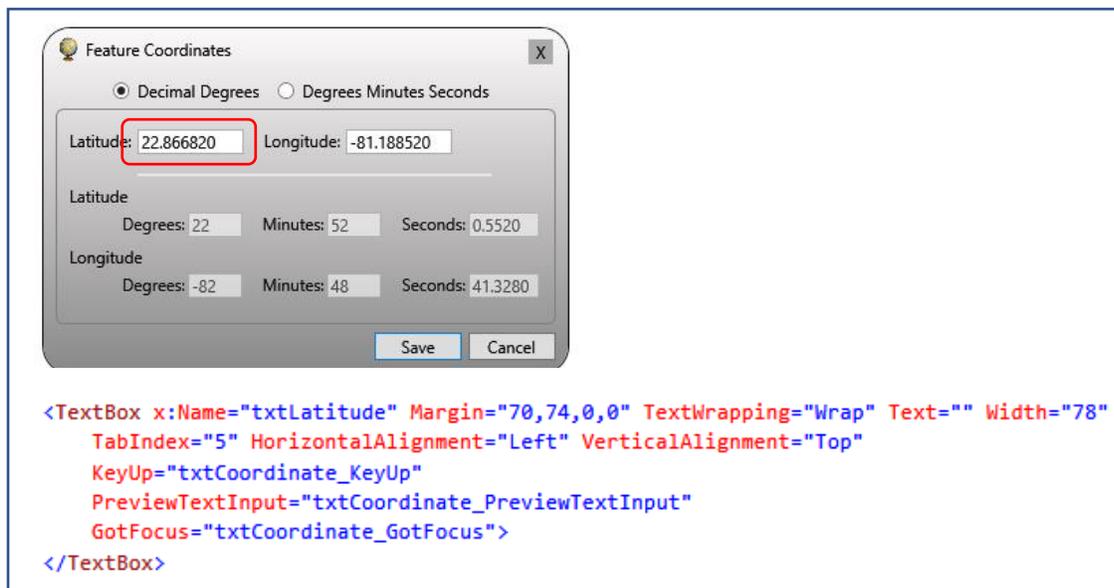
Figure 3.7: The form used to edit the coordinates for a map feature. The location, format, and actions for the input field circled in red are defined by the XAML snippet shown below the screenshot.

Interacting with the system

A mantra of historical research is to understand the *people*, *places*, and *events* recorded in the texts. Therefore, a robust information management system was required to not only record these individual components but allow them to be tightly linked within the underlying data structures and visually navigable from within the user interface. The essential components of the system allow the researcher to organize and manage libraries of source materials, enter transcriptions, extract content, establish links between people, places and events, and publish the results to external data services.

The following sections provide a synopsis of the major functional components of the user interface. It is important to note here that this is by no means a complete discussion of the capabilities of the system. For the sake of brevity, many of the features used to streamline data entry and facilitate navigation are omitted from the discussion and the information presented below refers only to the primary document management interface (Figure 3.8) along with individual windows used to manage and link People, Places, and Events (Figure 3.9). Please refer to the video clips in Appendix D for a more comprehensive presentation of the user interface and system functionality.
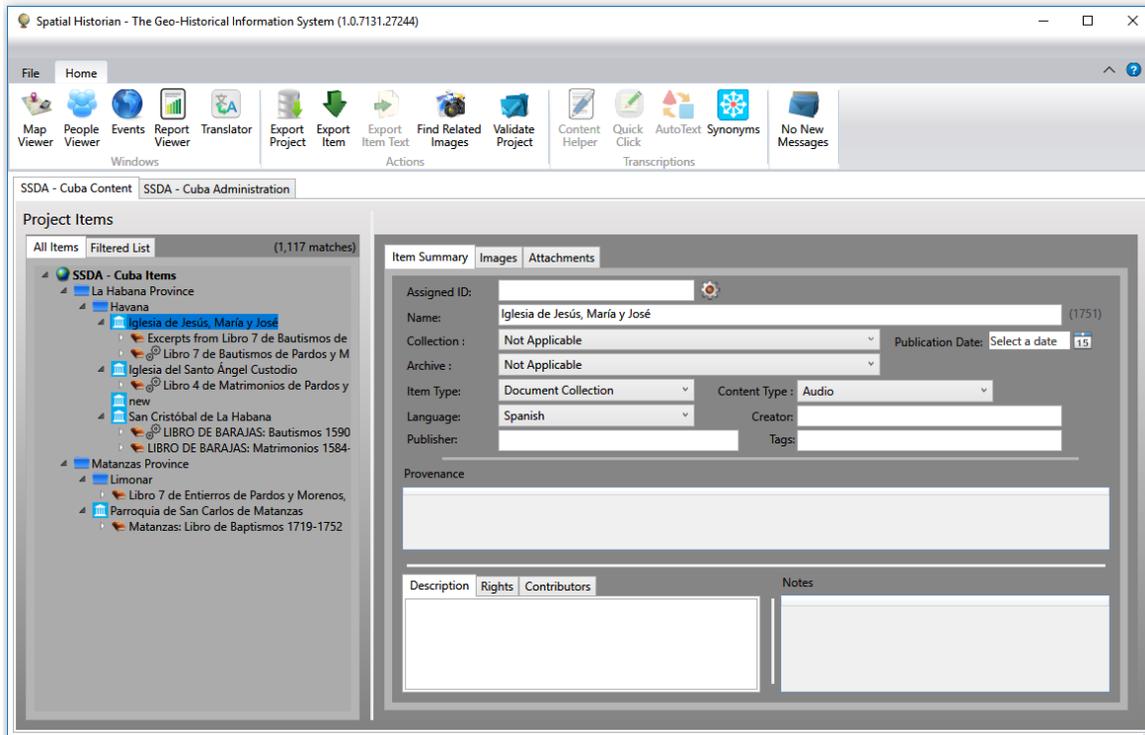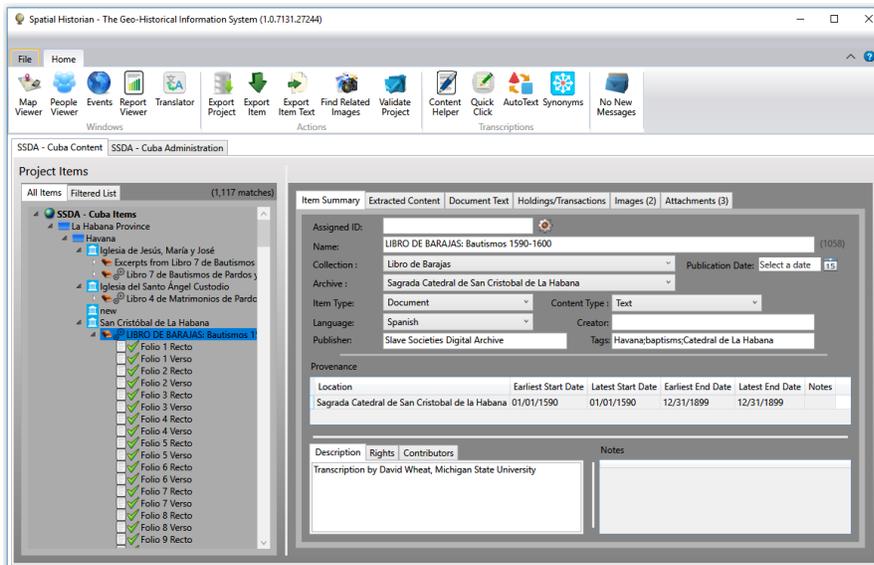
Figure 3.8: Project specific document management interface. The document hierarchy is customizable to support individual project needs.
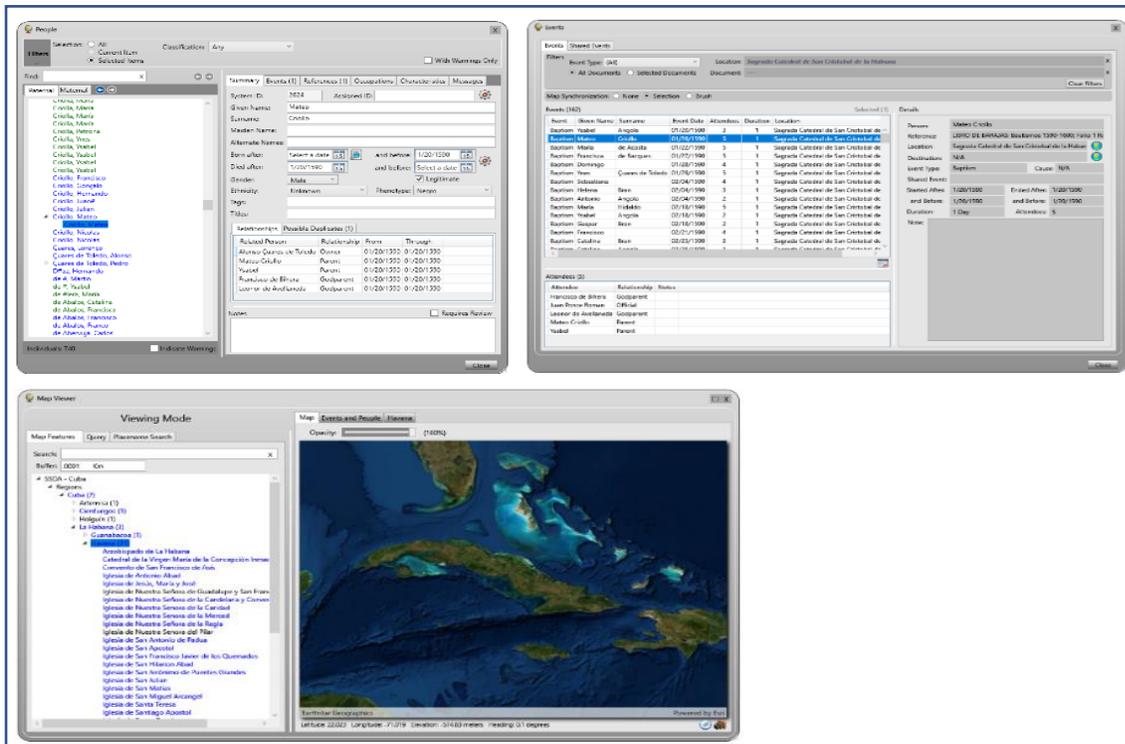
Figure 3.9: Primary windows for managing People, Places, and Events.

Document and Transcription Management

When working on a project in the Spatial Historian, the user is initially presented with the document management interface as shown in Figure 3.10. The hierarchy tree on the left of the screen is fully customizable based on the organizational preference of the research project. In the example shown in Figure 3.10 the documents are organized by the provinces, cities, and churches in which the individual document volumes were found. The volumes themselves are then divided into individual folio sides. This organization mimics the hierarchy of the document images in the example projects collection, however the hierarchy can be configured based on the specific needs of each individual project.

Figure 3.10: Project specific document management interface. The document hierarchy is customizable to support individual project organization needs.

Details about the individual items in the document hierarchy are managed using the tabs located on the right side of the window, with the available maintenance options varying based on which type of item in the tree is then currently selected. In the example in Figure 3.10, the baptism register named *Libro de Barajas: Bautismos 1590-1600* is selected in the hierarchy. On the *Item Summary* tab to the right the volume is defined as a *Document*, indicating that it is a standalone

physical item, and its primary language is Spanish. The *Provenance* section shows the locations and times where the document has been located over its life. This information is used to construct metadata for export and curation, and for aiding in associating physical locations to the events recorded in the volume. The additional fields on this tab are additional metadata used as well for export and curation.

Because documents, as well as other items in the hierarchy, do not exist in isolation, the system allows the user to add attachments to items that may provide additional background and context. In the example in Figure 3.11, an image of the cathedral where the document is located has been attached along with two supporting MS Word documents. The system supports the attachment of essentially any type of content and allows the attachments to be ingested and stored within the system or to be referenced by specifying a URL.
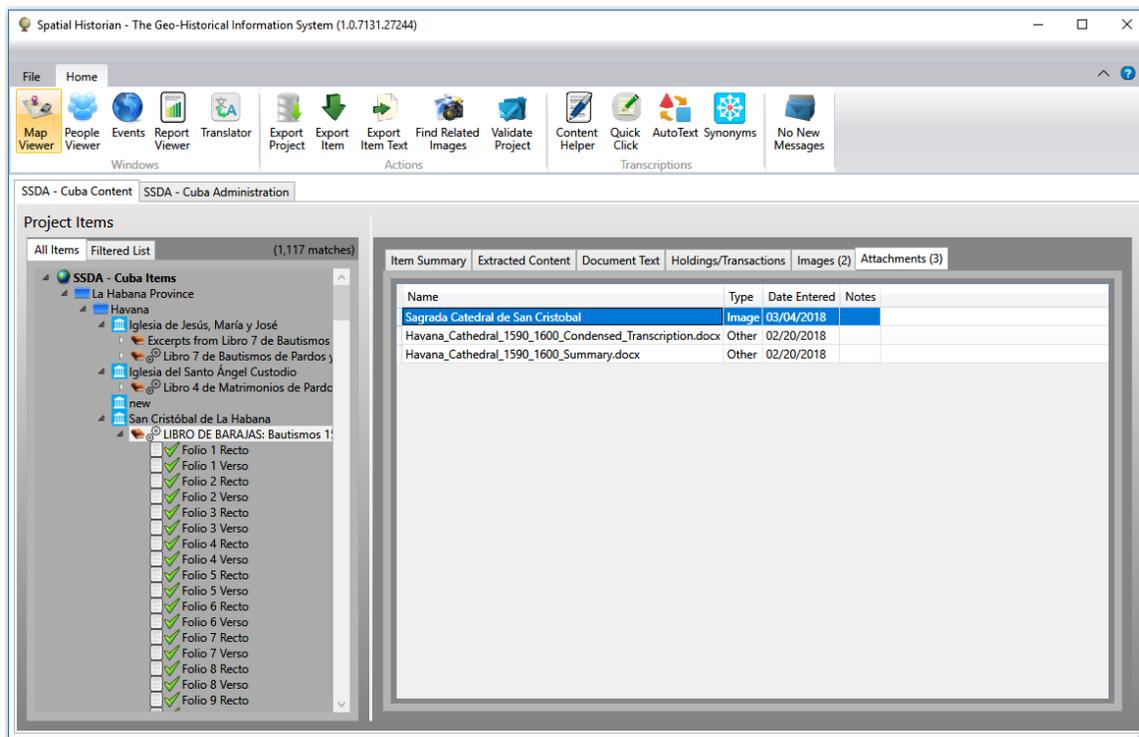
Figure 3.11: Attachments linked to a selected document.

In addition to attachments, the system allows a user to upload images of the selected item. The example in Figure 3.12 shows an image of Folio 1 Verso of Libro de Barajas: Bautismos 1590-1600. Once added, the image can then be used as a visual reference when transcribing the corresponding page. As an aid for transcription, the user can zoom and pan the image as well as open it for viewing in a separate popup window.
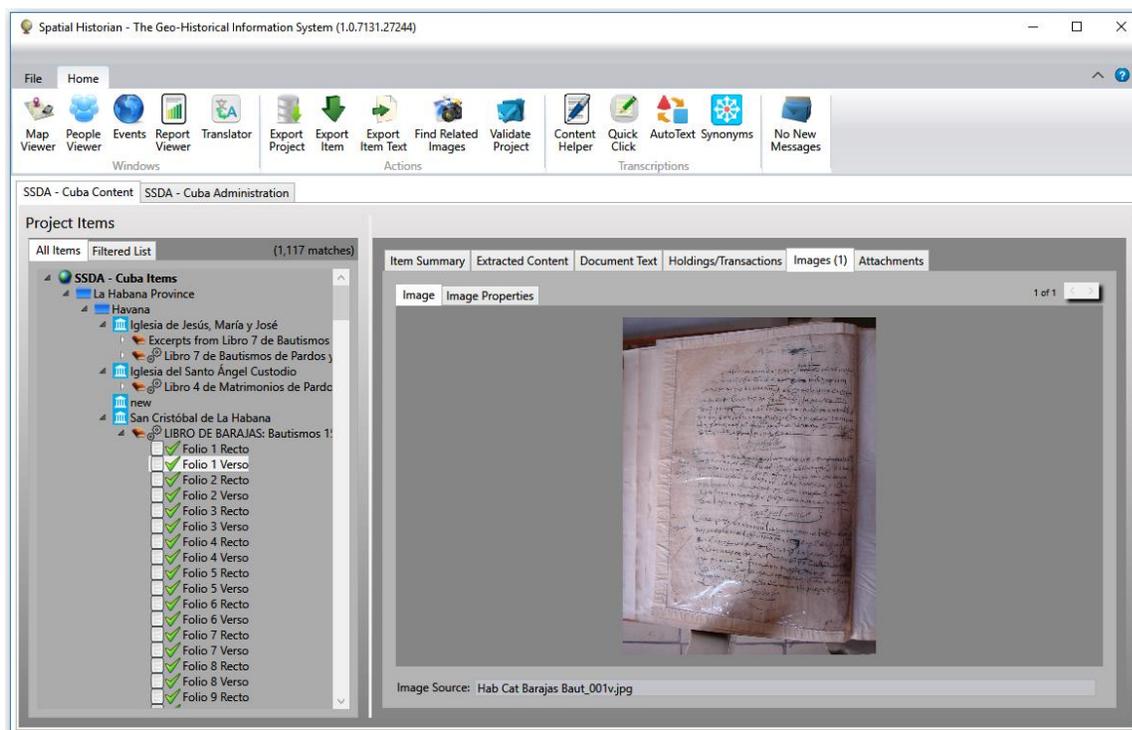
Figure 3.12: An image of the selected folio used for curation and for supporting the transcription process.

When a folio is added to the hierarchy, the transcription itself can be entered and manipulated using the *Document Text* tab as shown in Figure 3.13. This tab supports several functions including entry of the raw transcription, a normalized version (that is one in which spelling variants have been converted to standardized forms), and a translated version. From within the *Document Text* tab, the user would usually enter the raw transcription directly and then instruct the system to generate both the normalized and translated versions.
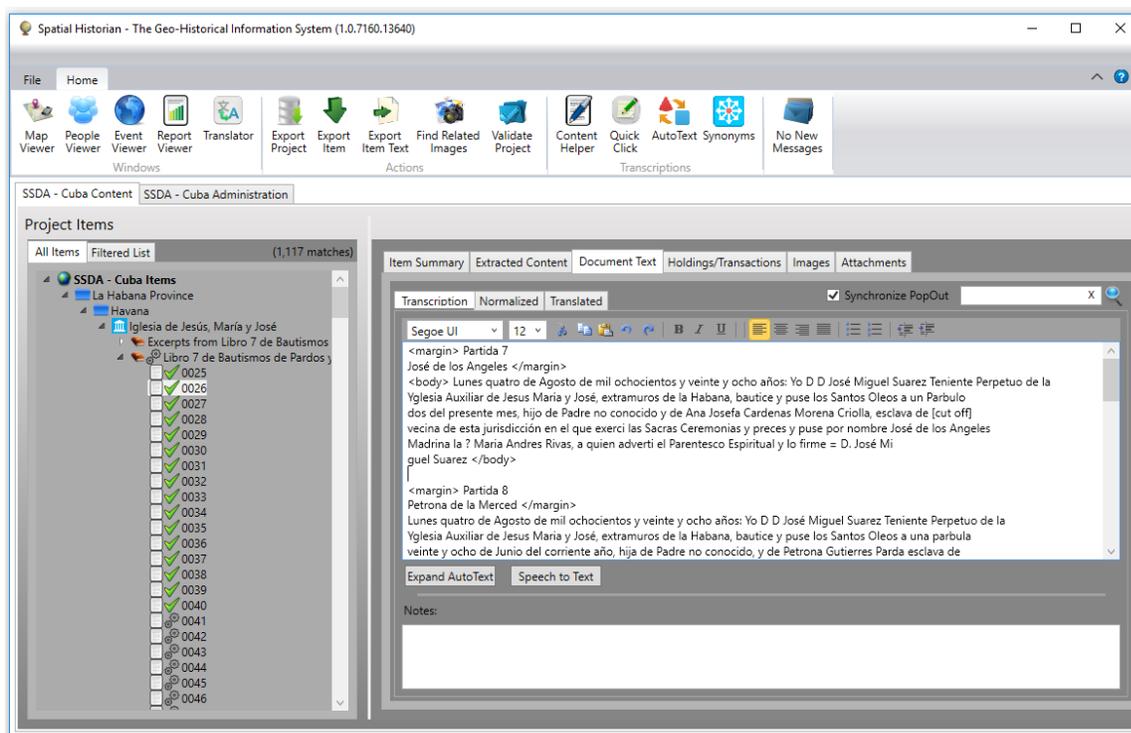
Figure 3.13: Folio text as initially transcribed.

Content Extraction using Synonym Matching

Content extraction is the process of identifying and recording details about the people, places, events, and relationships recorded in the texts. As with the transcription process itself, content extraction is a time-consuming and error prone manual process. To help mitigate the potential for errors and to help speed the process, several automated functions have been designed into the system. One of the functions is the automated identification of the names of people and places with the aid of a synonym matching feature.

The example entry recording Mateo Criollo's baptism that is detailed at the beginning of Chapter 1 of this dissertation contains important information, but it also offers an instructive example of one of the challenges encountered when analyzing historical texts which is the spelling

variations that can exist for the same word. The example entry identifies *Alonso Çuares de Toledo* as a slave owner and *Fancisco de Bihera* as Mateo's godfather though the name Alonso was written "Al⁰" and Francisco was written "Fran^co". These types of idiosyncratic abbreviations are common throughout many historical documents. In addition, there are often differing spellings for the same name such as "Fernández", "Fernandez" and "Fernándes", where phonetically similar letters are interchanged, or diacritics are applied based on the preference of the writer. Accounting for synonyms is important when trying to identify the names of individuals and place, and when producing normalized versions of the texts. The Spatial Historian addresses this issue using customizable synonym lists that are configurable based on standard languages, or project specific dialect or toponym requirements. The lists are made up of *basewords*, which are the standard versions of words, and their associated spelling *variants* and abbreviations. Therefore, to account for the abbreviation Fran^co for the name Francisco, the list contains a *baseword* entry spelled "Francisco" and a *variant* spelled "Fran^co" as shown in Figure 3.14 (By convention the '.' in the variant is used to indicate that the following characters are superscripts). Note also in the figure, that more than one *variant* can be defined for each *baseword*, and that *basewords* can be assembled hierarchically. Therefore, in the case of the name "Fernández" there is a variant spelled "Fernandes" that itself has the variant "fern.s".

Figure 3.14: Synonym hierarchy showing optional spellings and abbreviations for Fernández and Francisco.

In some instances, synonyms can be examined to determine if a person referenced in a text has been identified previously. For example, *Alonso Çuares de Toledo* may be recorded as a slave owner in a subsequent entry in the *LIBRO DE BARAJAS: Bautismos 1590-1600* or some other volume, but rather than using the abbreviation "Al°" the name may be spelled "Alonso". When a request is made to add a new entry, the system scans the list of existing names looking for name matches for all possible combinations of *basewords* and *variants*. If matches are found, the user is notified and can then investigate further to determine if the new transcription refers to the previously identified person or to a new person with the same name.

The synonym list is also used for a process called *normalization*. This is a process whereby the text, as initially transcribed, is converted to a *normalized*, or standardized, format and all

spelling *variants* found in a text are converted to their *baseword* equivalents. Therefore, in the normalization process the spelling "Al.o" in the initial transcription would be converted to "Alonso". Thus, the line "*Mateo Criollo y Ysabel Criolla su mujer, negros de Al.o Çuares de Toledo*" would be converted to "*Mateo Criollo y Ysabel Criolla su mujer, negros de Alonso Çuares de Toledo*" in the normalized version. Creating and storing a normalized version of the texts provides a consistent rendering of the contents that can be valuable for further examination and yet the original spelling transcription with its higher fidelity to the source material is retained.

The Spatial Historian also employs a master names list containing over 30,000 given names and surnames in several languages. Synonyms are used in conjunction with this list as an aid to extract content by allowing a user to instruct the system to scan transcriptions to identify any names that may be embedded. This process produces two results; the first highlighting pieces of text that might be interpreted as a part of a name, and the second is a list of people already recorded in the system. An example of the results of such a scan is shown in Figure 3.15. In the example, all names found in the text are highlighted in blue and a list of 21 potentially matching people, already recorded in the system, is displayed.
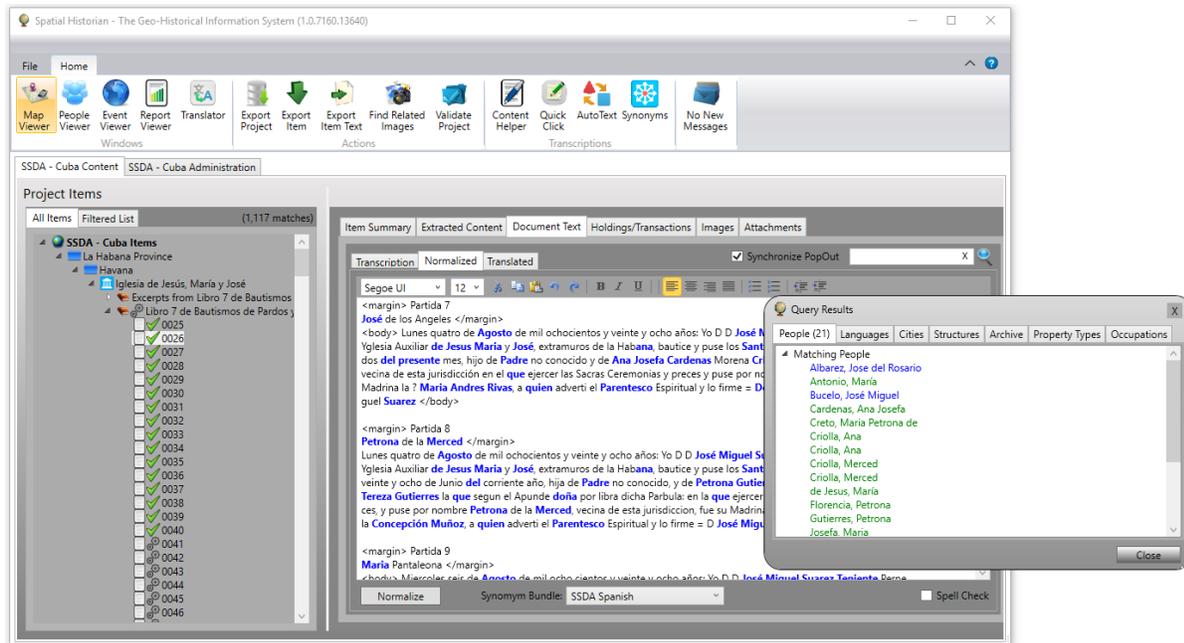
Figure 3.15: Results of scanning the normalized text of a folio. All names found in the text are highlighted and a list of potentially matching people already recorded in the system is displayed.

People

A fundamental premise of this system is to support the content management of the extracted historical text along with their related details. The following sections summarize how the People, Event, and Map windows are used to manage and analyze these data.

The People window is the primary user interface for managing details related to individuals (Figure 3.16). This interface is used to create, update, merge, and delete record entries. The left side of the window contains a hierarchical list of individuals with tabs that allow viewing by either paternal or maternal lineage. When a person is selected from within the hierarchy, the related details are displayed in the set of tabs on the right. The summary tab shown in Figure 3.16 contains basic demographic information along with a list of individuals who are related to the selected person. The demographic data is updated by entering new details directly into the fields on the form while the references and other information is updated using content specific pop-up forms.

In the Figure 3.16, *Mateo Criollo* is selected in the list of people and his information is shown in the detail section. Referring to the previous discussion of Mateo, the *Summary* tab indicates that he was born prior to January 20th, 1590, he was classified as a *negro*, and was considered *legitimate*. Also shown are his relationships to his slaveholder, his parents, and his godparents. Figure 3.17 shows, on the left, the list of known events associated with Mateo along with the people in attendance, and on the right, the list of known references.

An important feature of the system is that the user can directly navigate to any of the people, places or events from any of these listings. Thus, to navigate to the details for *Alonso Çuares de Toledo* in this example, the user would right-click on his name in the list of relationships and select the "Go to Person" option from the pop-up context menu. Similarly, to navigate to a document, the user can right-click on the reference and select "Go to Document". The ability to navigate between people, places, and events is available from any listing in the system.
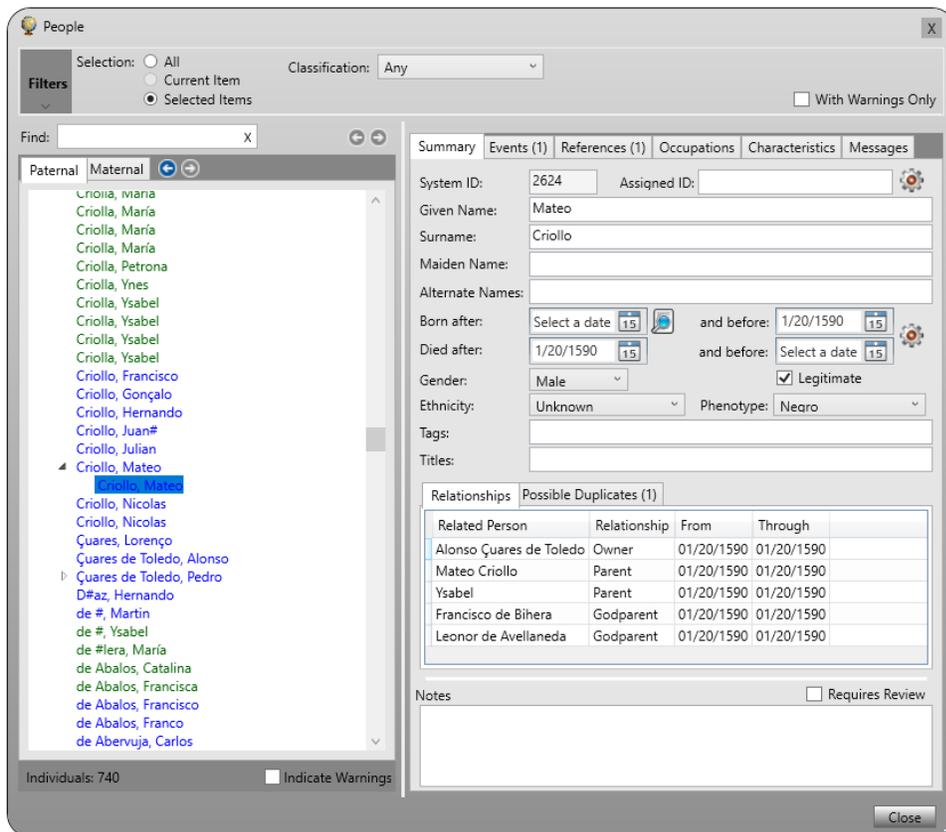
Figure 3.16: The People window with a hierarchical list of individuals on the left and details for the selected individual on the right. In this example, the information for Mateo Criollo is displayed.
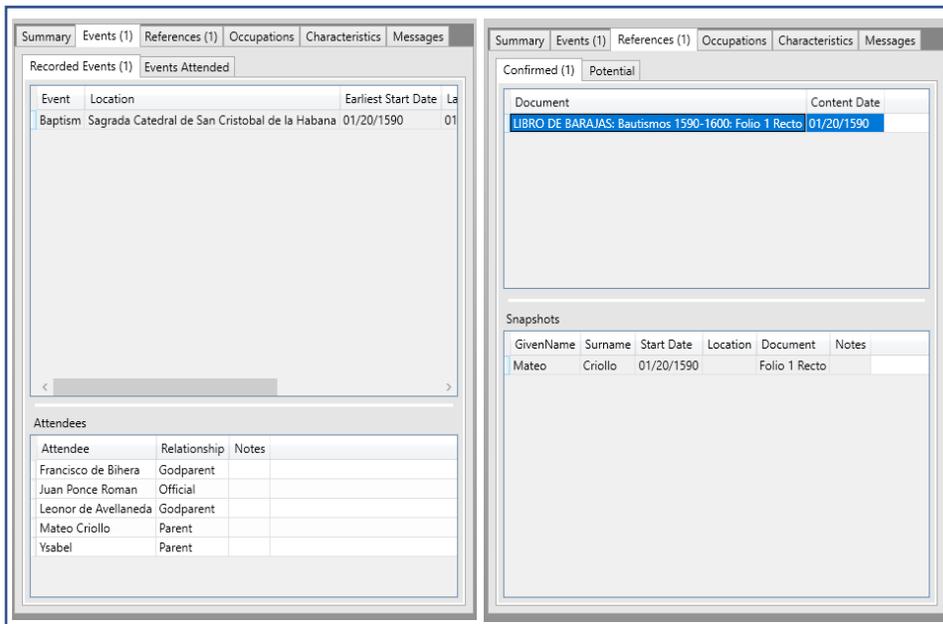
Figure 3.17: Events and References for Mateo Criollo

Events

The *Events* window is the primary interface for viewing and managing events specific to individuals as well as *Shared Events* that are events that affect more than one individual. Figure 3.18 shows an example of the *Events* window with the baptism event for Mateo Criollo selected. The bottom left of the window lists the attendees at the baptism and the right portion show details of the event. From this window the user can navigate from the list of events directly to the reference document or the person, and from the Attendees list to the individual attendees' details. Also, because location information can be stored for each event, the user can instruct the system to navigate to the location of the event on the Map window by clicking on the globe icon in the Details section of the window.
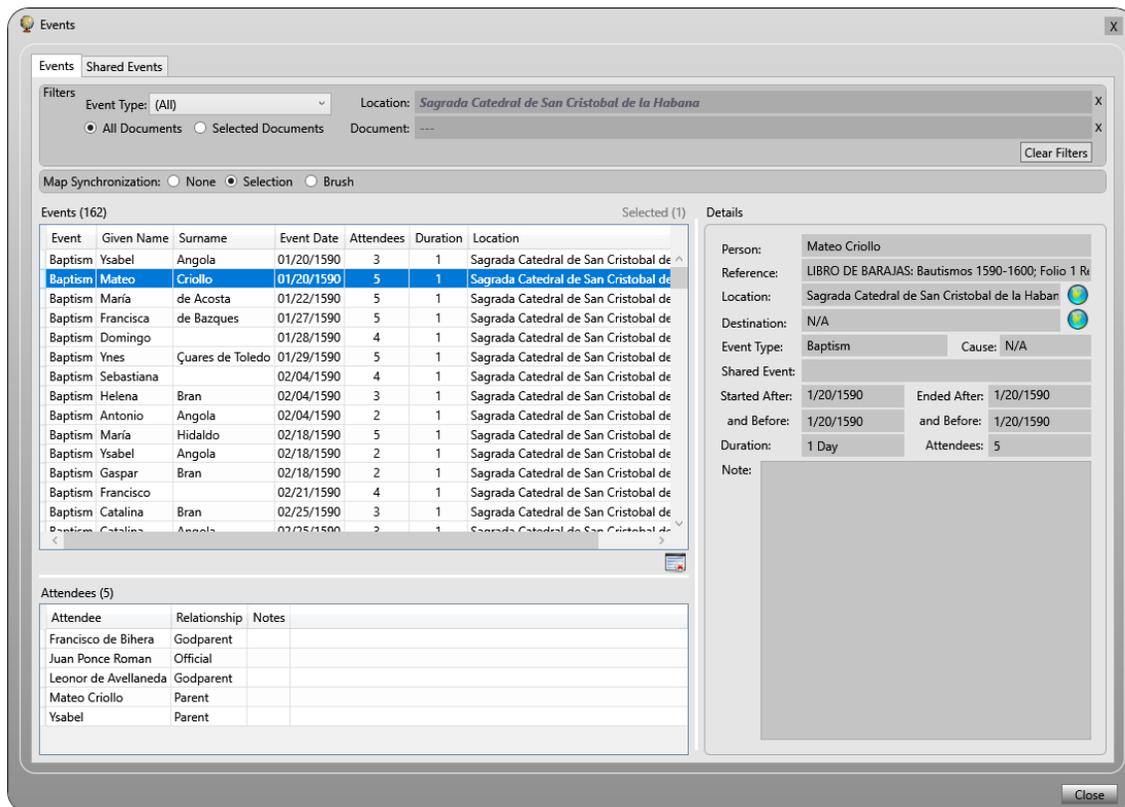
Figure 3.18: Event listing window with the baptismal event for Mateo Criollo selected. The bottom left of the window lists the attendees at the baptism and the right portion displays details of the event.

Shared Events affecting more than one individual have both a temporal and a spatial extent. The Bubonic Plague that began in Europe in the 1340's, or the voyage of a slave ship carrying its crew and cargo of enslaved persons would be example event. In general, *Shared Events* result in life events for individuals such that the exemplar plague may result in both a death and a burial event for a person, while the arrival of a slave ship at a port may result in a *Disembarkation* event and likely a subsequent *Sale* event for an enslaved individual. Shared events provide important linkages in the social networks of those affected by event even if individuals never met. *Shared Events* can also supply implicit information about the affected individuals. As such, the records of enslaved baptisms that occurred within a short period of time after the arrival of a slave ship may indicate that the individuals being baptized arrived on the ship and therefore came from the region where the ship was loaded with human cargo. These implicit connections, while they require large quantities of records in order to validate the hypothesis, demonstrate the types of exploratory questions that can be generated using the data collected and managed by the system.

Figure 3.19: Listing of Shared events showing the type and duration of each.

Mapping

The Map Viewer window is the primary interface for displaying the places and spatio-temporal locations of events recorded in the historical documents. The interface consists of a project specific hierarchy of places on the left and an interactive 3D map on the right (Figure 3.20). For the SSDA Cuba project a hierarchy of Region, Country, Province, City, and Church is defined. The spatial extent or point location for each item in the hierarchy can be defined if known or named placeholders can be created for toponyms whose location is unknown. Figure 3.20 shows an example plot with the locations for a subset of churches in the city of Havana plotted.
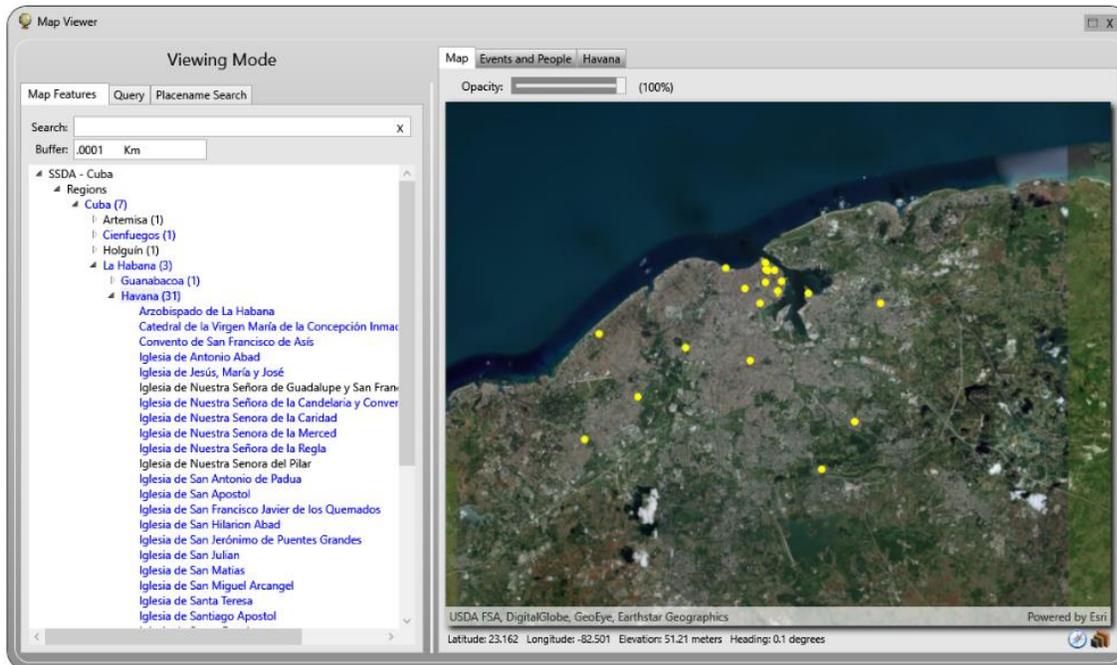
Figure 3.20: Map Viewer with churches located in the city of Havana identified.

This simple display of the point location of churches can be enhanced by embedding 3D models within the viewer. When the user zooms in to a specific location, as in Figure 3.21, the symbology is updated to display the 3D representation. In this example, stock models are shown, though models of actual structures produced from photogrammetry, lidar, or other sources can be used when available.
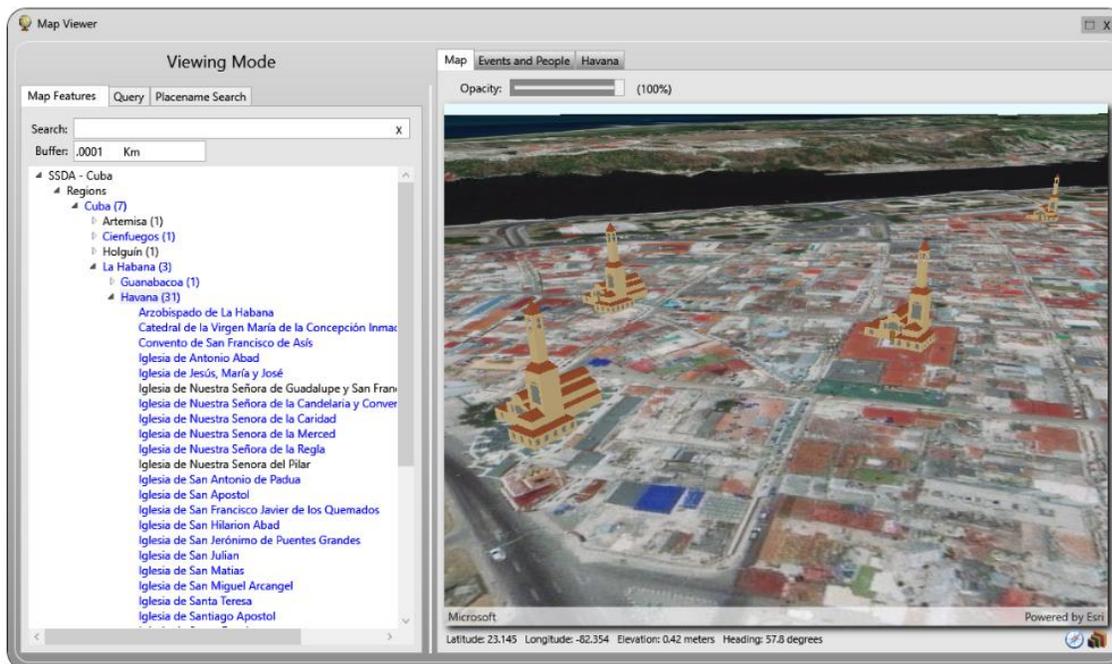
Figure 3.21: Map symbology using 3D models.

Utility Functions

Beyond the interface functionality required to process the texts and manage the extracted content, the Spatial Historian provides several utility functions designed as general aids for the system user. These include a feature allowing the user to search the web for images based on keywords; a window that performs basic text translation; and a window that displays general project statistics and project validation results. There is also an embedded message board that allows system users to leave messages for one another and a report viewer that renders custom designed reports. Access to these utilities is through toolbar items on the main document interface window.

Image Search

The Image Search functionality is built on Microsoft Cognitive Services (Microsoft, 2018) images search service. Searching is initiated by clicking the *Find Related Images* toolbar item as

shown in Figure 3.22. When the tool is activated the user is presented with the screen as in Figure 3.23. This window allows the user to enter any set of comma or semicolon separated search keywords. By default, the field is automatically filled with the tags defined for currently selected items in the document hierarchy. In the example, the tag values "Havana; baptisms; Cathedral de la Habana" have been automatically entered by the system. These tags can be edited as desired and additional parameters defining the number of results to return, the Safe Search setting, and the scope of the search can be set.
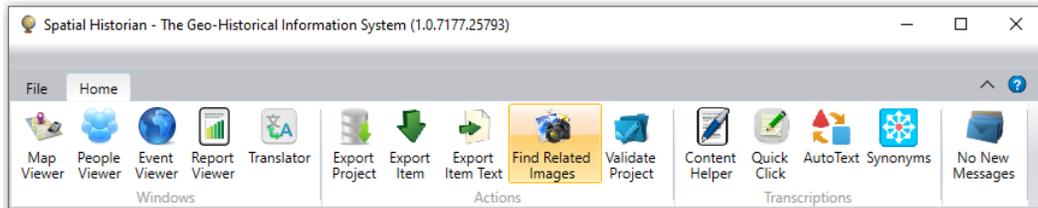

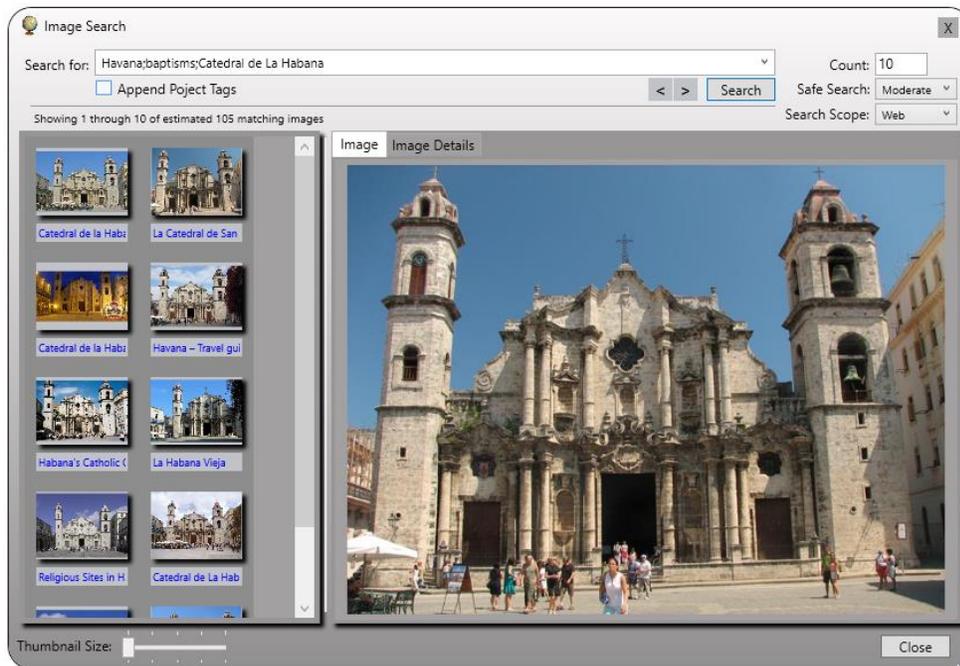Figure 3.22: Image search tool highlighted in orange.

Figure 3.23: Keyword based image search feature.

Clicking the *Search* button initiates a search with the results displayed as thumbnail images on the left of the window. The slider at the bottom left allows the user to adjust the size of the

thumbnails as desired. Clicking on a thumbnail image causes the full resolution version to be downloaded from the source and displayed in a larger format on the right. Right-clicking on the large image presents the user with a menu that allows the image to be rotated or attached to the currently selected item as either a linked or an embedded attachment.

In addition to the rendering itself, the image metadata is also available by selecting the *Image Details* tab (Figure 3.24). This tab displays metadata extracted directly from the image itself. This included information about the image source, the camera, the camera settings, and, if GPS is enabled, the latitude, longitude, and altitude of the camera when the image was taken. When the coordinates are available the *Plot* button is enabled. When clicked this button opens the Map Viewer and zooms to the corresponding location on the map.
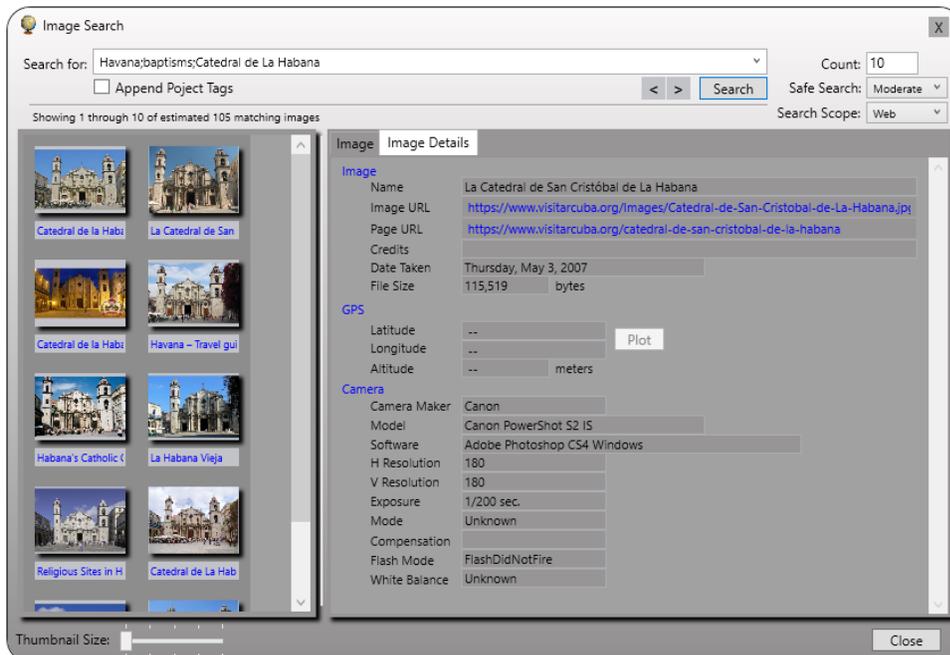

Figure 3.24: Metadata extracted from the image file. For images taken using a GPS enabled camera the latitude and longitude are populated. When coordinates are available, the user can click the Plot button to show the location in the Map Viewer.
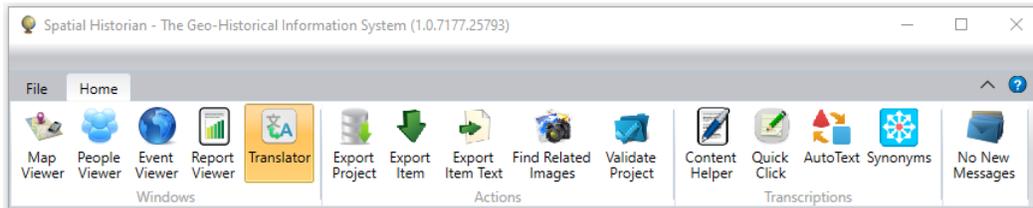
Translation

Figure 3.25: Translator tool highlighted in orange.

The Spatial Historian includes an intuitive interface that allows the user to translate a piece of text between any of 64 available languages. The translation is initiated by selecting the source and target languages from the corresponding drop-down lists, entering or pasting text into the *Source Text* field, and clicking the *Translate* button. When clicked, the source text and language specifications are sent to the Microsoft Azure *Translator Text API* (Micorsoft, 2018) cloud service with the results of the process displayed in the *Translation* field. This functionality has proven viable for short snippets of text. Longer portions of text with more complex content, however, are often misrepresented and therefore only an approximation can be expected.
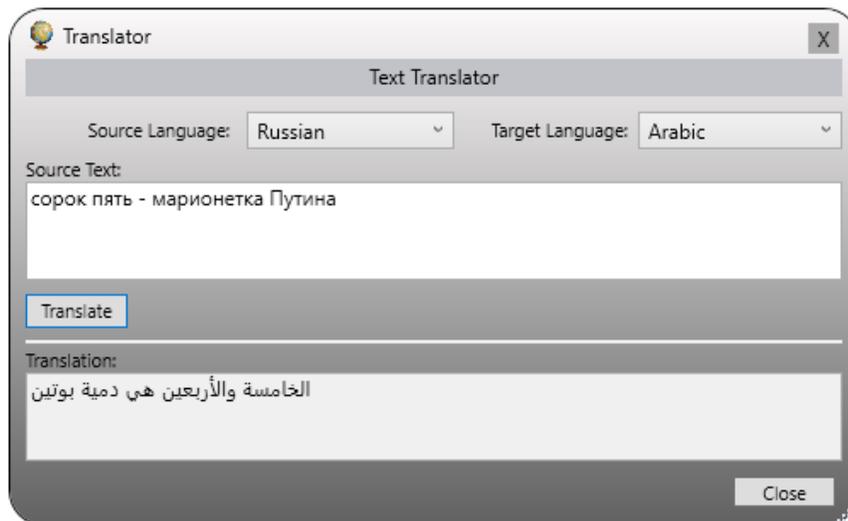

Figure 3.26: Translator window with text converted from Russian to Arabic.
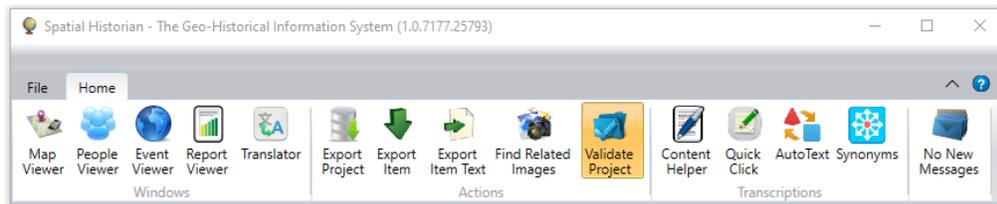
Project Summary and Validation

Figure 3.27: Project summary and validation tool highlighted in orange.

As the number of documents in a project grows and the amount of extracted information increases in step, it can be useful to have a quick reference point to generate overall system and project statistics. The *Project Data Validation* window was created to provide general statistics as well as perform a series of data validation operations looking for potential data entry errors and data inconsistencies (Figure 3.28). The first tab, labelled *Statistics*, presents counts of the number of items in the document hierarchy, the number of Map Features, and the number of people extracted. The tab also shows counts of extracted events classified by event type.
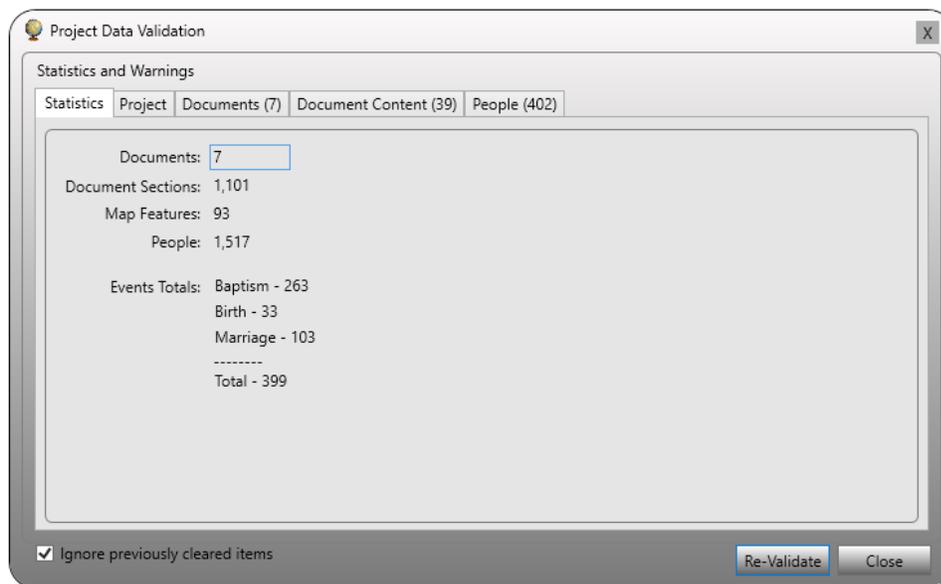

Figure 3.28: Summary statistics for a project.

The remaining tabs show the results of the validation for the overall project, including documents, document items, and people. The validation process and corresponding results are

specific for each grouping. Figure 3.29 shows the validation results related to documents while Figure 3.30 shows results of the validation process for people. The validation results are, however, only indications of potential problems with the data and they in no way affect or block any other functionality of the system. The results are presented solely as a starting point for further investigation. The system facilitates investigation by allowing the user to right-click on any of the result entries using a popup menu that allows quick navigation to the related item as show in Figure 3.30. In this example, clicking *Go To Person* will open the People Window and position the selection to *Asencion Suares*.
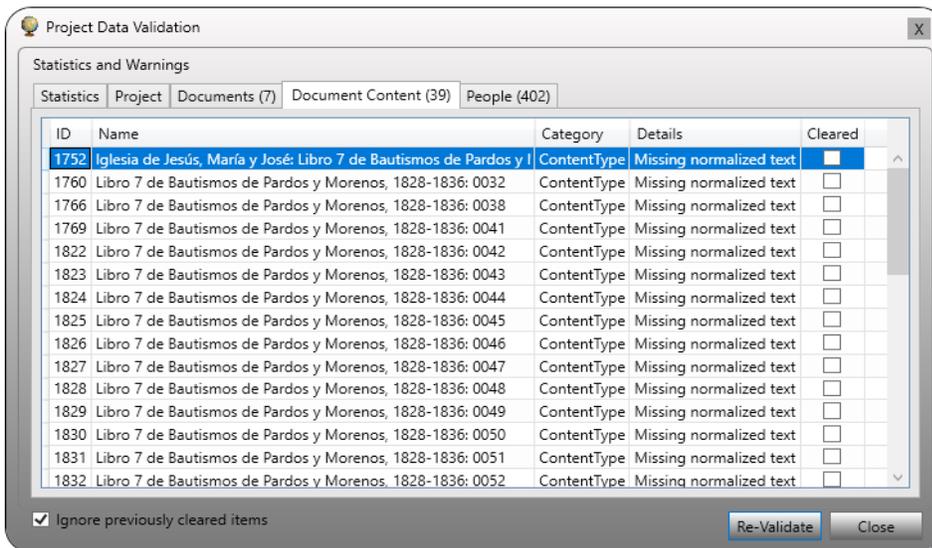
Figure 3.29: Warnings related to documents. These generally involve missing metadata.
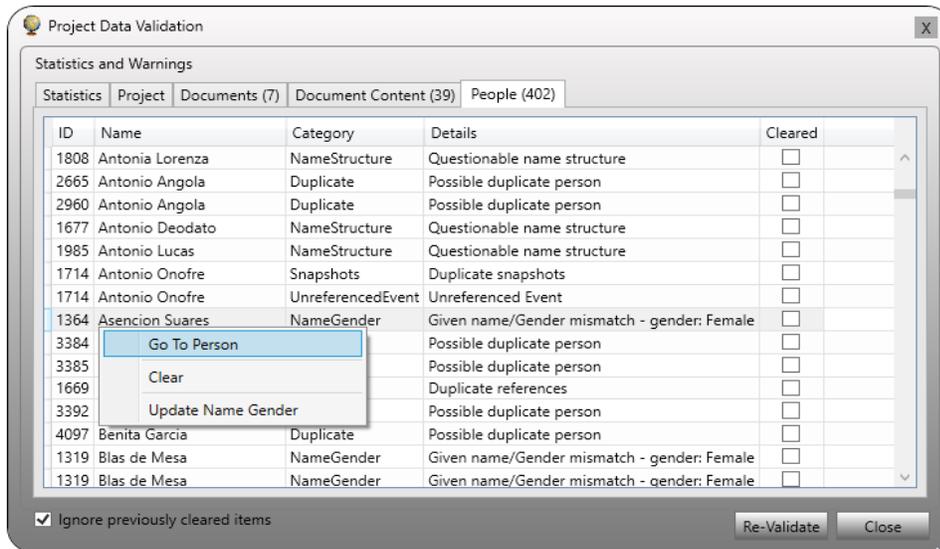
Figure 3.30: Listing of warnings related to people.

Message Board
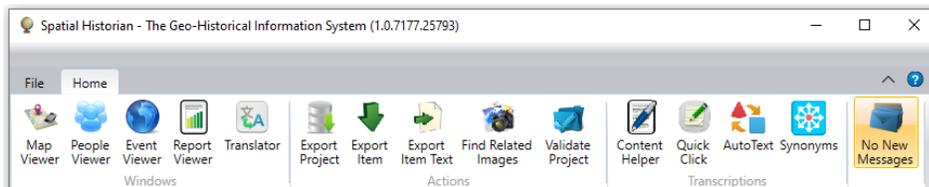
Figure 3.31: Message board tool highlighted in orange.
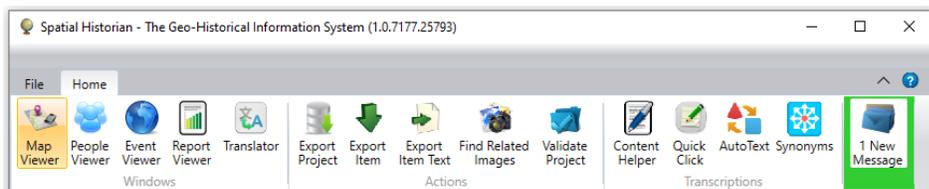
Figure 3.32: The Message Board tool highlighted in green with the text indicates the presence of a new message.

The Message Board within the Spatial Historian is a lightweight messaging module that allows users to send and receive messages to one another. It is intended to be an easily accessible means of sending communications related to projects or other content managed within the system. The message interface, which is located under the User Settings tab on the initial system form, consists of a list of inbound and outbound messages for the current user (Figure 3.33). New

messages are created by clicking the envelop icon located above the message list. Clicking the envelope opens a form that allows the user to select recipients, create a subject line, and enter a message (Figure 3.34). Clicking *Send* on this form places the message in recipient inboxes with a copy placed in the senders S*ent Mail* list. Entries in the lists can be flagged as read or can be deleted as appropriate by right-clicking and selecting the corresponding option from the popup menu.

Figure 3.33: The internal message board that allows users to communicate within the system.
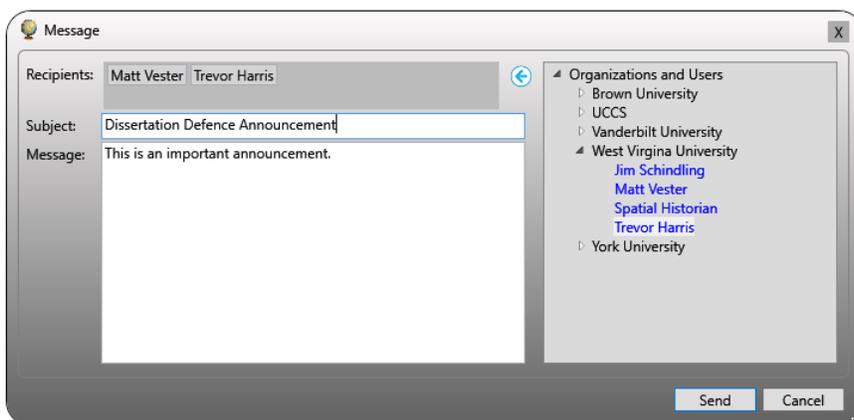
Figure 3.34: Composing a message.

Reporting
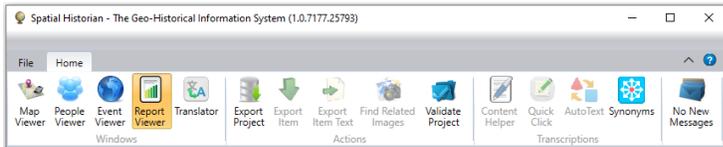
Figure 3.35: The Report Viewer tool highlighted in orange

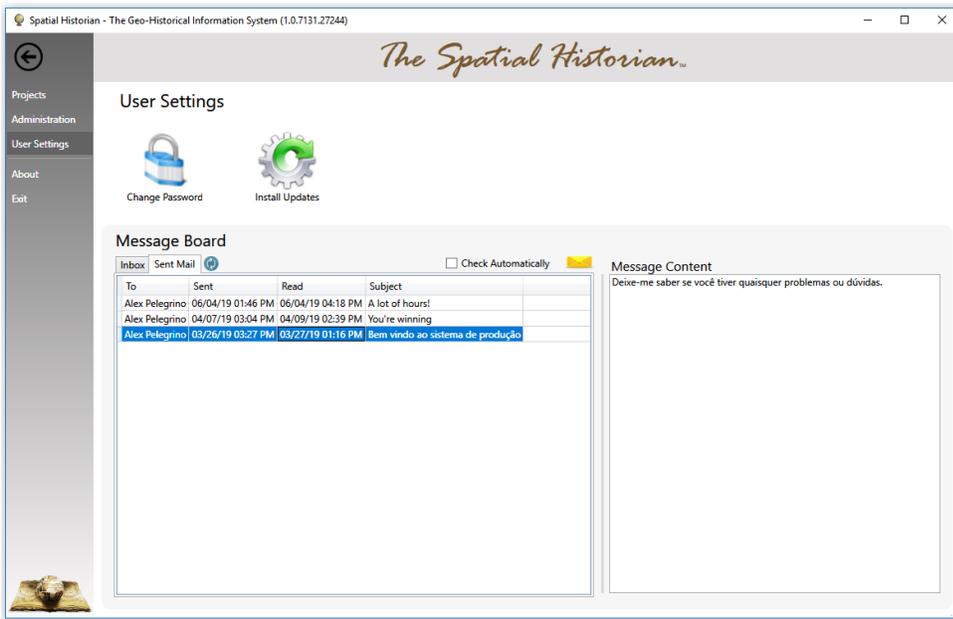Because the Spatial Historian stores data in a SQL database, any user with appropriate database permissions and a knowledge of one of the many ad-hoc SQL query tools can query the database and generate reports on demand. However, as a starting point, the system offers a reporting interface based on a set of reports hosted in Microsoft SQL Server Reporting Services (SSRS). SSRS is a server-based report generator that allows users without direct database access to execute predefined reports. The results can be viewed or exported in several formats, such as Excel, for further offline use.

Clicking the Report Viewer tool opens the window shown in Figure 3.36. Once opened, the user selects the desired report from the dropdown list in the upper left and clicks the *Load Report* button. The list of available reports is based on a configurable system parameter set by the system administrator. This parameter specifies the URL for the SSIS server. At runtime, the server uses this URL to provide the system with its list of reports that are used to populate a selection list.

The SSIS technology allows reports to be designed with interactive functionality. In the example shown in Figure 3.36, the user can click the '+' and '-' boxes to dynamically expand the report sections. Options are also available to embed hyperlinks to related content such as web hosted versions of document transcriptions, additional research materials, or images of structures or artifacts.

Figure 3.36: Custom window rendering reports hosted on a cloud-based Reporting Services instance. This example reports the enslavement status of godparents of baptized slaves.

Virtual Reality

Accessing 3D Virtual Reality models is implemented in the Spatial Historian using appropriately configured attachments. For example, in Figure 3.37, the first of two attachments to the church *Iglesia del Santo Ángel Custodio* is set to reference a custom VR application containing a model of the structure. This approach allows any number of VR models to be linked to any type of object within the system including documents, archives, people, and map features. This approach also allows any or all VR mechanisms discussed below to be implemented simultaneously thereby offering the broadest range of support.
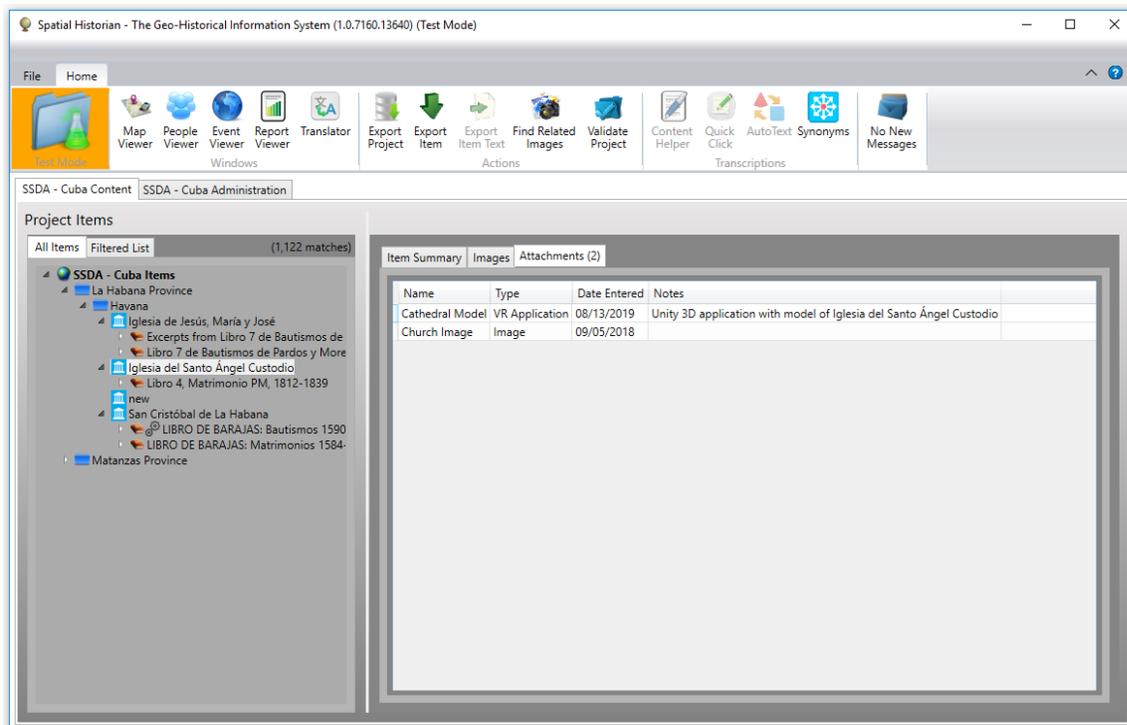
Figure 3.37: The first attachment for Iglesia del Santo Ángel Custodio is configured to launch a custom VR application.

There are four different mechanisms for viewing and interacting with VR attachments within the Spatial Historian:

1) As custom VR applications located on the user's network

2) As VR applications on sites such as Valve Corporations Steam VR

3) As hosted models on sites such as Sketchfab

4) Using Spatial Historian's custom 3D model viewer.

The first of these options assumes that a custom application has been developed using some VR development tool such as Unity 3D. Using this approach, a VR software developer would embed one or more 3D models directly into an application. When the application is executed, the user would be able to interact with the models based on the options and navigation techniques as implemented by the developer. Figure 3.38 shows how an attachment using this approach would

be configured. Note that the *Source URI* field contains a reference to an executable program located on a network accessible drive and the *Content Type* is set to VR Application. An advantage of this configuration is that the executable is directly under the control of the user or the system administrator and is therefore likely to be reliably accessible. If the VR application is developed for multi-user access, this option also opens the possibility of the collaborative analysis of artifacts by allowing multiple researchers to view, manipulate, and discuss the objects together within a virtual environment. The disadvantage is that, because it is on the local network, it will likely not be available if the Spatial Historian is run outside of the local network.
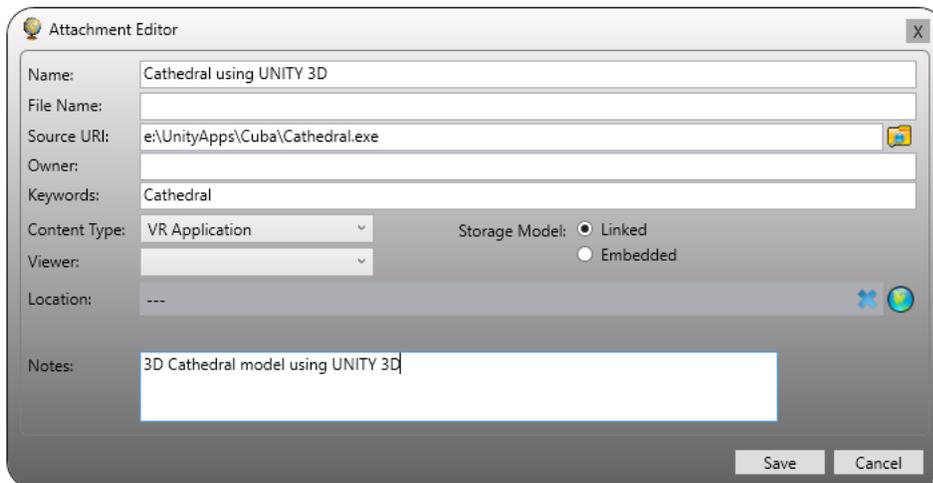
Figure 3.38: Attachment configured to launch a custom Unity 3D application.

The second mechanism is based on the Steam VR virtual reality application environment developed by Valve Corporation (Figure 3.39). The Steam VR environment allows developers to upload VR applications to be hosted on their web portal. Uploaded VR applications can be accessed over the internet, and the process can be automated using the Valve Software Development Kit (SDK) (Valve Corporation, 2019). For example, the command shown in Figure 3.39 will start the application with the identifier number 1020. The advantage of this approach is that the VR model can be accessed from over the internet allowing it to be available within the

Spatial Historian regardless of where the user is located. The disadvantage of this approach is that it requires the Steam VR SDK to be installed on the user's machine.
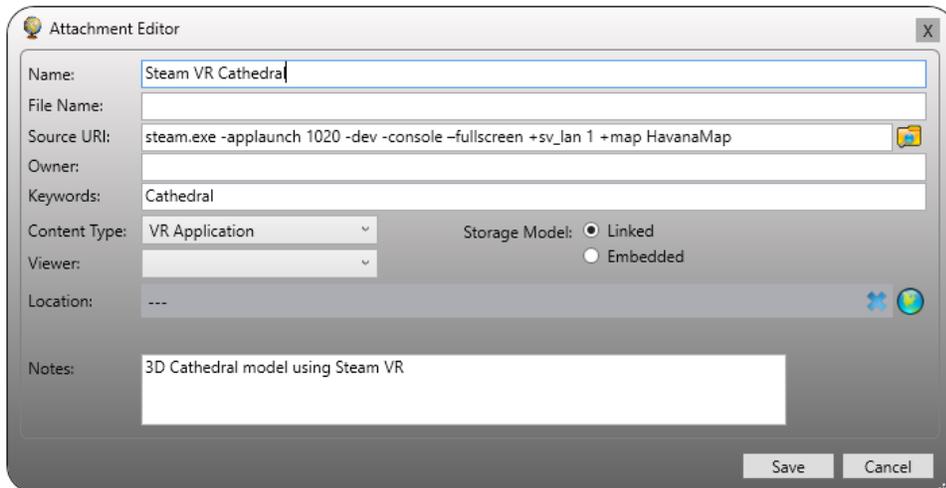
Figure 3.39: Attachment configured to launch an application using Steam VR.

The third mechanism accesses 3D models published on sites such as Sketchfab (Figure 3.40). The Sketchfab site allows developers to upload 3D models to be hosted and viewed on their web portal. Uploaded models can be accessed over the internet without the need to load any software on the user's machine. For example, the example command in Figure 3.40 loads a model of the Notre Dame cathedral. As with the previous mechanism, 3D models can be accessed over the internet allowing them to be available from within the Spatial Historian regardless of where the user is located.
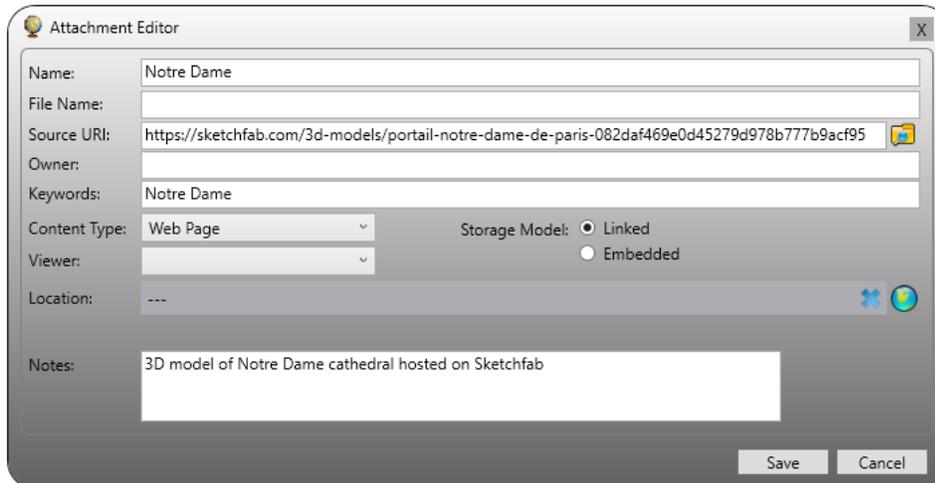
Figure 3.40: Attachment configured to launch an application using Sketchfab.

The fourth option uses a custom Unity 3D application that takes advantage of the software's ability to dynamically load object models at runtime. This option is enabled by setting the *Content Type* to *3D Model Viewer* and the *File Name* to the name of an asset bundle containing the 3D model to be loaded. For example, in Figure 3.41, the attachment is set to run the *3D Model Viewer* using the model contained in the file *LevalloisLithicsBundle*. The advantage of this approach is that it allows any number of models to be viewed without the need to create a separate application for each. Also, because the *3D Model Viewer* is installed as part of the Spatial Historian, it is available to all users. A disadvantage is that the viewer, although functional, is currently in beta test mode. Also, the format of Unity 3D asset bundles is proprietary and require that objects in standard formats be manually imported prior to use.
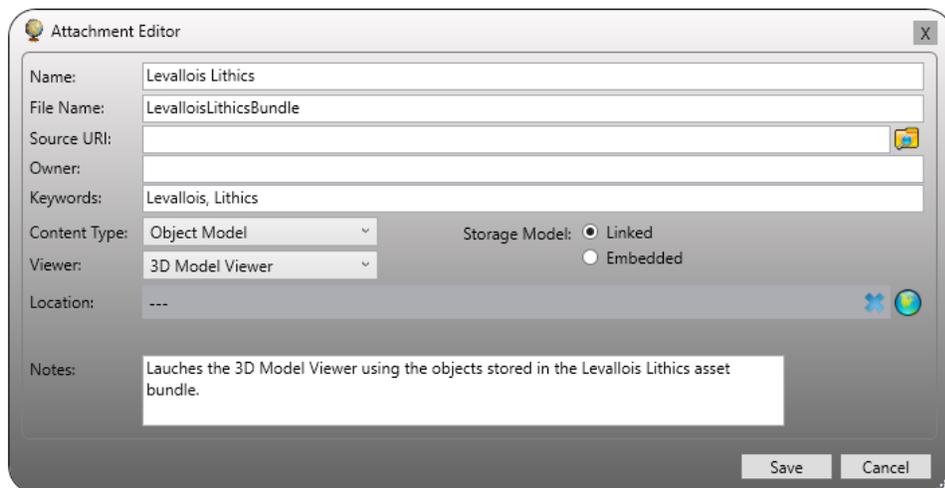
Figure 3.41: Attachment configured to view an 3D model using the Spatial Historian Object Viewer.

Curation – Data Export

An important role for any software system is to address mundane tasks, and to make them as transparent and unobtrusive as possible. For historical research projects this includes the capture and recording of project specific data along with the metadata required for data interchange and curation. The management of metadata and derived data produced during research is an often-neglected aspect of humanities research, either because they are not of immediate concern to the research objective or because the mechanisms for their management are not well understood. The Spatial Historian addresses this deficit in two ways. The first approach is to provide the user with jargon-free interfaces for entering metadata as a procedural aspect of using the system. For example, by allowing a user to enter keywords and ownership data related to documents, the system captures metadata elements that can be used to produce METS compliant metadata. A second approach is through the support of a drag-and-drop assignment of spatial location information related to documents and events. This process records not only the names of where documents have been located or where events have occurred, but also Open Geospatial Consortium (OGC) compliant spatial details (Open Geospatial Consortium, 1999; Open Geospatial

Consortium, 2010; Open Geospatial Consortium, 2011) and SQL/MM (Stolze, 2003). The choice of Microsoft SQL Server for data storage was made, in part, because of its conformance to OGC and SQL/MM spatial data standards. By storing spatial data in SQL Server geography data types, all metadata required to support applicable spatial data curation requirements are inherently recorded and available for export. This includes support for *geography* data types which support ellipsoidal (round-earth) data, GPS latitude and longitude coordinates, and standard Spatial Reference Identifiers (SRIDs) (International Association of Oil and Gas Producers, 2019). By implementing this storage mechanism, the system is capable of producing Spatial Data Transfer Standard (SDTS) (Bleakly D. R., 2002), Federal Geographic Data Committee (FGDC) (Metadata Ad Hoc Working Group, 1998) and ISO 19115 compliant metadata (McGarva, Morris, & Janee, 2009).

The programming model used for the Spatial Historian supports the creation of any number of data export objects inheriting from a base export class. A base class named *ExporterBase* (ExporterBase.cs) defines standard logic for Excel format and comma separated value, or CSV, file creation and writing. It also includes virtual methods that can be implemented to create files in other formats, such as XML or json. As of the time of writing, two derived classes have been implemented that generate files in the format specified by the *Enslaved* project at Michigan State University. This is a project specific format based on the requirements defined for the Slave Societies Digital Archive case study described in chapter 5.

The data export process is initiated from the main toolbar on the Spatial Historian user interface as shown in Figure 3.42. The user has the option of exporting all data for a project or exporting only the data related to selected documents. In either case, the export process is initiated by clicking the *Export Project* tool which opens the Exporter window (Figure 3.43) and allows the

user to select the file format to be generated. When the *Export* button is clicked the user is prompted for the name of the folder in which to place the output files. Depending on the format selected, one or more files are created and placed in the folder. The resulting files can then be made available for transport to the appropriate external system based on the pertinent specifications. For example, some external systems may require that the files be consolidated into a single .zip (compressed) file or that they be left as individual files.
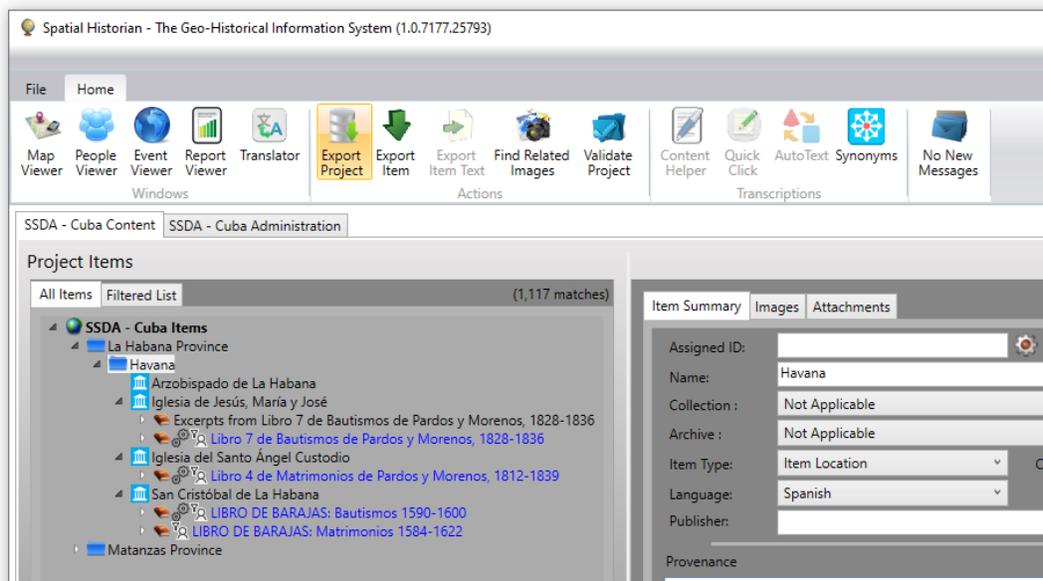
Figure 3.42: Exporting project related contents is initiated from the main document management interface by clicking the Export Project toolbar item, highlighted in orange.
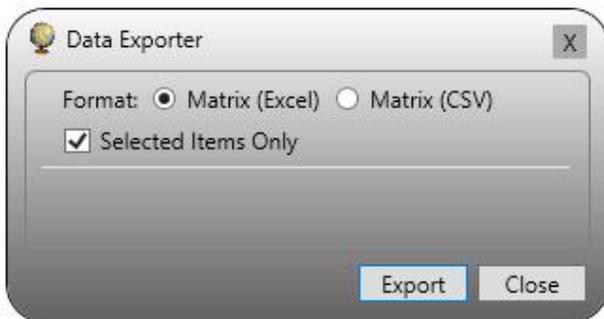
Figure 3.43: The Data Exporter window that allows the user to specify the export file format and initiate the process.

CHAPTER 4 - Applying the System: Sixteenth-Century St. Vincent

To fully test the design of the system as a viable technology and as an aid to historical research, two case studies were undertaken; each differing in the makeup of their available source materials and each with distinct research objectives. The following two chapters discuss the distinctive features of these case studies, how these features informed the development of the final system, and in some instances, how the system exposed additional avenues for historical investigation.

This first case study is based on a research project examining ways in which the familial and social structure the sixteenth-century town of St. Vincent, Italy are reflected in the city's notarial records and other documents recorded during the period. This is an ongoing research project under the direction of Dr. Matthew Vester of West Virginia University. The goal of the St. Vincent project is to "spatialize kinship relations within a rural mountain community during the early modern period, in order to compare family dynamics within and between rural families (in small towns and villages) with those in other social settings (in cities, amongst nobles and elites, and even within and between ruling dynasties)" (Appendix B) (Vester, 2019). Two sets of primary source materials for the area were used for this case study comprising a set of sixteenth-century notarial records and a collection of nineteenth-century maps commissioned at the time by the Italian government.

Several characteristics of the St. Vincent project made it a particularly applicable case study for testing the technology and functionality of the Spatial Historian and its applicability to spatio-historical research. First, the collection of nineteenth century maps offered the opportunity

to evaluate the processing required to ingest hand-drawn maps for use in a modern GIS. Second, the source materials offer the opportunity to evaluate the integration of 16th century historical texts with the maps and current satellite imagery, through the use of toponyms as the connecting thread. Third, because of the use of inconsistent place names in the notarial records and the variations in spelling between the notarial records and the maps, the St. Vincent case study presented a challenge to the system to disambiguate place names, and to use person names to develop deep family genealogies. Finally, the digital spatialization of property holdings required the linking of transaction details recorded in the historical notarial registers to the current landscape.

The Source Materials - Texts

The available text materials offer examples of documents representative of the region during the 16ᵗʰ century. The notarial records from the *Archivio Notarile Distrettuale di Aosta* for the city of St. Vincent document among other events, the transfer of property and land between individuals. Similar to the baptismal records from Cuba, the notarial records of St. Vincent provide an indication of where and when the document was executed, and the names of the individuals present. In addition, the records provide explicit details about the location and size of the property being transferred, and the relationships of the parties involved. In an entry spanning two pages recorded December 1, 1565, for example, a piece of land of one *setier* in size made up of a field used to grow grapes, was assigned to Jean Nodyn by Vincent Seris for the price of ten florins. The land was located at a place known by the toponym Crest d'Orchires within the territory of Clappeya. From this individual transaction, a social relationship between Nodyn and Seris can be inferred, although it is not possible to know if this was their only interaction or just one of many. Given the spatial information, it is also possible to assemble a picture where grapes may have been grown in the region, the value of that land, and an indication of the size of parcel that might be used for grape production. However, as with other historical documents, additional value lies in the aggregation of large numbers of records over extended periods of time. In this way additional references to vineyards can be used to create a map of regional land use and shifting values and uses over time. Additionally, transactions involving these parties may be used to develop a picture of land-holding strategies and of family growth or contraction.

Several factors combine to challenge the analysis of the St. Vincent texts. The documents themselves are unstructured and written in a unique paleographic style that does not lend itself to interpretation using automated Optical Character Recognition (OCR) software (Figure 4.1). A

further complicating factor is that the documents were written in a variation of Franco-Provençal called Valdôtain which, if automated, would require the development of a customized translation procedure. The extraction of spatial information is made more difficult because place names used in the records are often descriptive colloquial terms rather than formal administrative names as would be common today.
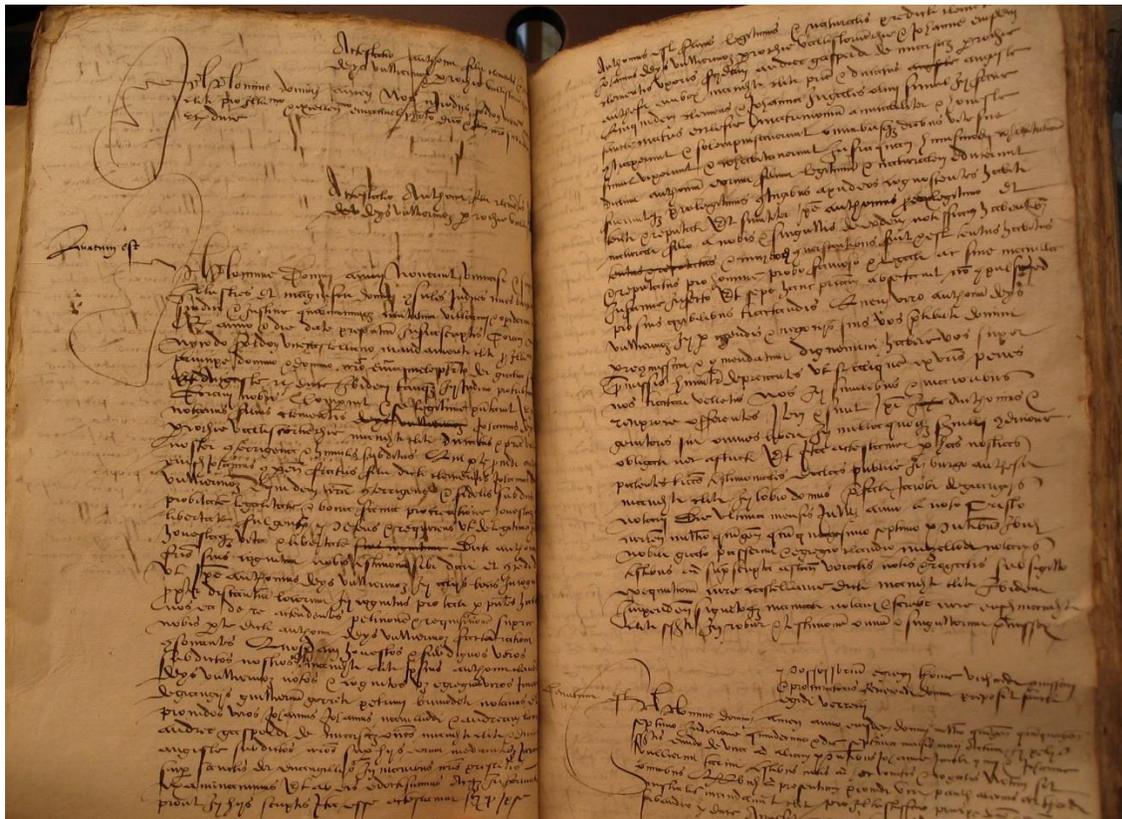
Figure 4.1: Sample St. Vincent notarial register entry

Assigning locations to the people, places and events mentioned in the texts is made more difficult because there are no existing maps for the region for the period under study. The maps closest to that time period were produced for the Italian government in the late 19th century to support tax assessment and collection. In the intervening period the spelling of many place names had changed or were lost or replaced completely. Fortunately, the maps do provide geographic detail at the level of individual land parcels and building structures, and are accompanied by a

record of the toponyms in use at the time of the map creation. The names, although often different in spelling from the sixteenth-century record forms, have proven to be durable over time and allow many of the toponymic references in the text to be accurately located.

The source materials available for this study test the capability of the system to capture and process data from unstructured historical materials; support collaboration between researchers possibly located at different geographical locations; provide a spatial context to historical records; publish materials and research results for both private and public access; and assign pertinent metadata to support export to a digitally curated archive. The St. Vincent case study is also significant in that it provides an opportunity to examine ways in which the many shortcomings of the original prototype system have been addressed and the system enhanced beyond a technology testbed with the goal of becoming a universal fully functioning research tool.

The Materials – Maps

In the 1500s maps became important tools for extending monarchical power and establishing the discourse of control (Smail, 1999). Where the written declarations of holdings in tax surveys could be ambiguous, the location of properties on a map left little room for hedging and helped to ensure that the property was identified. The earliest maps of St. Vincent were created between the years 1893 and 1897 (Regione Autonoma ceila Calle D'Aosta, 2007). This set of maps includes an overview map at 1:10,000 scale along with 67 detailed maps at scales of either 1:500 or 1:1,000. An initial examination of the individual maps shows that they contain a great deal of detail and appear to have been based on very accurate surveys and map producers with well-developed cartographic skills. Individual parcels and structures are delineated and numbered. Each of the detailed maps is accompanied by an index of the territory (*frazione*) and toponym (*toponimo*)

in which each parcel was located. These documents, or *fogli*, indicate that the surveys themselves began ca 1892 and continued through 1895. The total area covered by the survey is approximately 41 square kilometers with a vertical relief of over 2,100 meters. The lowest point in the survey area is the Dora Baltea River at approximately 420 meters and the highest is the peak of Mount Zerbion at over 2,500 meters. The highest structures identified on the maps are located at over 1,600 meters in elevation. Interestingly, the large elevation gradient in this relatively small area is not noted on the maps themselves. Even though the people of St. Vincent relied on agriculture for subsistence and the town was located in a deep river valley in a mountainous region, there are no details related to elevation, the slope of the terrain, soil quality, or land cover. Additionally, the maps seem to offer no information that may have been useful for military, agricultural or even for travel planning. The focus of the survey seems to be exclusively on identifying and cataloging parcels of land for taxation purposes. Figure 4.2 shows an example of a source map. The label on the top left edge indicates that this map was produced at 1:1000 scale and had an original publication date of August 16, 1897. These black-and-white TIFF format raster images were produced from scans of the original hand-drawn maps.

These maps and their associated documents offer an important connection between the written records of the sixteenth century and the landscape as it exists today. The link from the maps to the earlier period is based on matching the map toponyms to the place names from the earlier notarial records. Linking these maps to the current landscape is also supported by the ability to georeference the maps to satellite and aerial imagery using GIS.
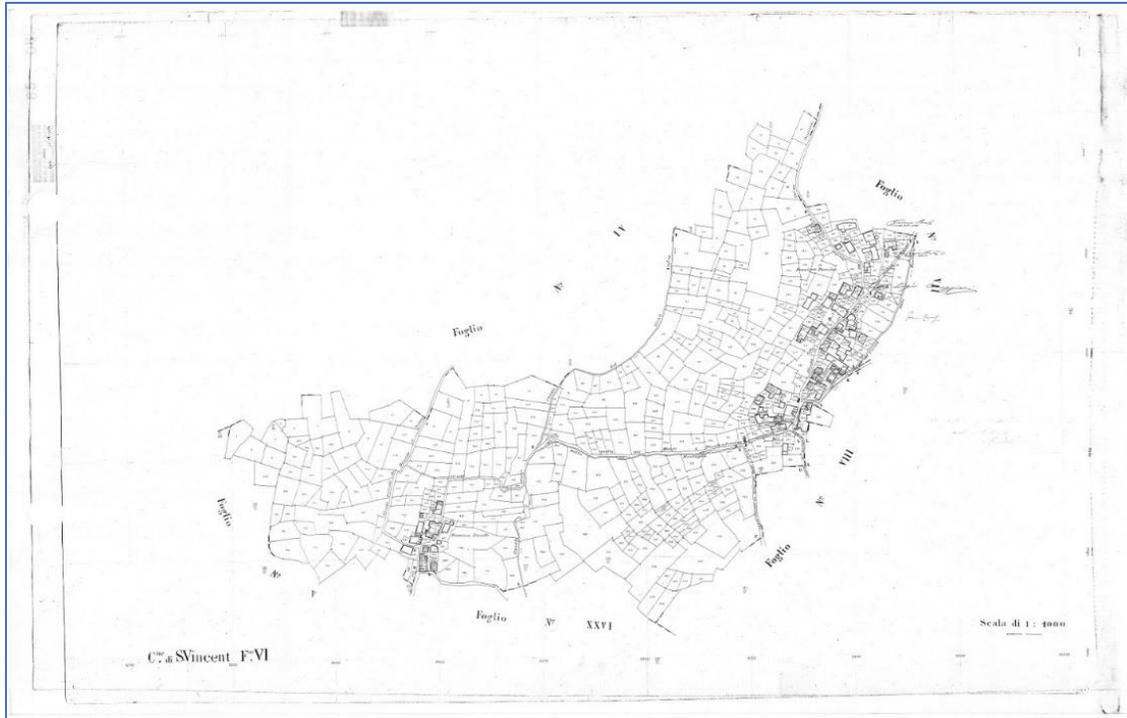
Figure 4.2: Map six of the collection showing the date that it was recorded, the scale, and an indication of the bordering maps.

Specific structures on the map can be identified based on their size and geometric shapes and because of the shading which is applied to the polygon boundary. The only indication of a functional use for any of the structures is a cross that appears within the boundaries of some polygons, implying perhaps a religious use. Of the 24 structures marked with crosses, three are named churches; St. Maurizio, St. Lorenzo, and St. Vincent, and four are smaller named oratories. Structures that may have held local secular significance such as mills or communal ovens are left un-named. This implies that the scope of the survey itself was for use beyond that of the local community. Another implication of the scope of interest of the map makers is that, along with the few churches that were labelled, transportation related structures were noted. By the time the survey was performed, a railway had been built through the Valle d'Aosta and a train station and four toll booths were deemed significant to be labeled.
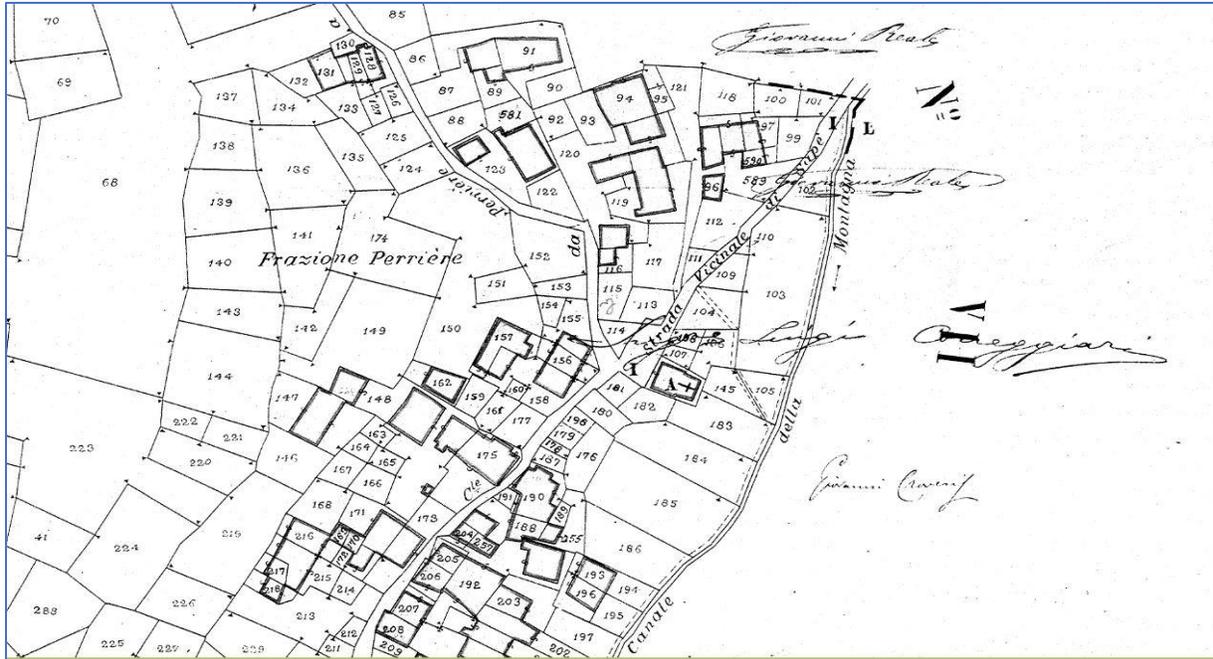
Figure 4.3: Detail of map 6. Land parcels are indicated by thin borders, and structures by thicker borders. Note also the labeling of streets and canals. The numbers within the polygons correspond to entries in the map ledgers. This map is unique among the set of 67 in that it is signed, although it is not known who the individuals were.

A close-up of a section of this same map (Figure 4.3) reveals some of the detail and visual indicators used to determine feature types such as roads and building structures. Land parcels are indicated by thin borders while structures are identified by thicker lines with internally shaded borders. A structure in the right central section of the map has a cross drawn in it possibly denoting a church or some other religious building although in this case it is not named. All of the parcels and many of the structures have numbers associated with them that correspond to an associated map index. The indices, in turn, indicate the territory and toponyms for the various numbered items. In Figure 4.3 it is possible to identify a grouping of structures as being part of *Frazione Perriere*, or the Village of Perriere, with the *Canale della Montagna*, or Canal of the Mountain, running north to south on its eastern edge. Two streets that intersect in the village are also labelled as the *Strada di Sape* and *Strada da Perriere*.

Application of the Spatial Historian

Dr. Vester articulates two specific difficulties he encountered in his research both of which deal with uncertainty in interpreting the data. One difficulty is in determining whether individuals identified in the various records who have the same name (and perhaps the same father's name) are in fact a single person, or whether they are different people. A second obstacle concerns the spatialization of family property and linking the spatial references in the sources to places that can be mapped in a GIS (Appendix B) (Vester, 2019). To help overcome these obstacles, the Spatial Historian aids in detecting possible duplicate individuals from within the extracted content and the property ownership denoted on the maps through the synthesis of both the texts and maps. The remainder of this chapter discusses the processing and user interface of the Spatial Historian as it establishes a list of people referenced in the texts and their relationships. This is followed by a discussion of the processing performed on the 19th century maps and how the resulting GIS data were integrated into the system and textual analysis.

The first of the two obstacles noted above is one that is common in the St. Vincent texts in that multiple individuals living within the same geographical area and at the same time might have the same names. This is problematic because when a person is referenced in a text, for example as a land owner, that person is recorded in the system along with an association to a piece of property at a given location. If a second individual with the same name is also identified as the owner of a different property it cannot be assumed that both entries refer to the same person who might own both properties. Additional complexity is introduced when individuals have similar or synonymous names rather than exactly the same name. For example, in the St. Vincent documents the name *Jacques* can be variously spelled as "*Jacques*" or "*Jaques*". Similarlly the surname "*Regis*" is

synonymous with "*du Rey*" and as a result, names recorded as "*Jacques Regis*", "*Jaques Regis*" and "*Jacques du Rey*" may, or may not, refer to the same person.

A step toward determining whether two individuals with the same names are the same person or not is to look at the names of their ancestors on the assumption that matching parent names would indicate that the individuals are likely to be the same. However, extending the comparison to look at the names of the ancestors can often turn up duplicates as well. In the St. Vincent documents, for example, there are two people with the name "*Jacques Regis*" and one name "*Jacques*", each with a father named "*Vincent*". It could be assumed that if sufficiently extensive family trees could be constructed the uniqueness of names could be determined. However, because of the limited numbers of available documents and limited resources to transcribe those documents, a family tree could end beyond only a generation or two.

Barring other contextual information or insight on the part of the researcher, the fallback approach for the St. Vincent project in these instances is to record separate individuals each with their own related property or properties. Then, if additional information comes to light that clearly identifies them as being the same person, they can be merged with their respective properties. This approach has its drawbacks especially as the number of duplicate names increases. Furthermore, the names may be placed under different branches in different family trees making identification even more difficult (Figure 4.4).

Because duplicate names are encountered frequently in the St. Vincent documents, the Spatial Historian provides an automated mechanism for detecting potential duplicates for any

individual. Automating the process, rather than requiring system users to manually record and manage possible duplicates, eliminates the need to create additional database structures and takes the burden off the researcher to recognize synonyms and remember possible other name associations. The first step in the Spatial Historian process looks for individuals in the database whose given names and surnames are the same as the person being evaluated. That comparison is then extended to look for synonymous versions of the names. The resulting list provides a baseline set of possible duplicates. This list is then enhanced by adding a flag to each entry indicating whether the names (and synonyms) of their ancestors also provide a match. An additional test is performed to calculate the spatial proximity between the potential duplicates and the person being evaluated. This propinquity is determined by examining the location of each individuals' event or events for which the individual was recorded. The result of this process is presented in the People Viewer as shown in Figures 4.4 and 4.5. In this example, five possible duplicates for the currently selected "*Jacques Regis*" are listed in the *Possible Duplicates* grid and the duplicate individuals are highlighted in the list of people on the left. Note that the grid contains all possible matches based on given name, surname, and synonyms. Check boxes are included in the listing to indicate whether the ancestral names on the maternal or paternal side also match to at least one generation. A popup menu on the grid allows the user to quickly jump to the record for any of the listed entries. If, upon further examination, it is determined that an entry in the list is in fact the same person, the context menu provides the option to merge the duplicate with the currently selected person (Figure 4.6).
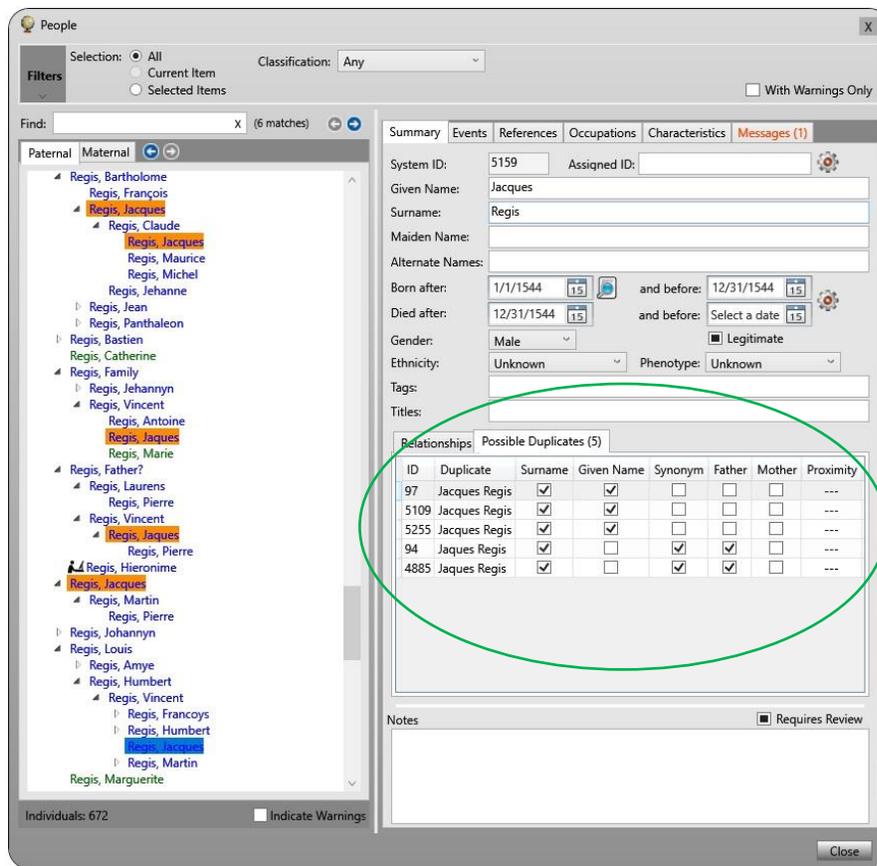
Figure 4.4: Identification of potential duplicates for Jacques Regis. Five possible matches are listed in the Possible Duplicates grid (circled in green) and the duplicate individuals are highlighted in orange in the hierarchy of people on the left.
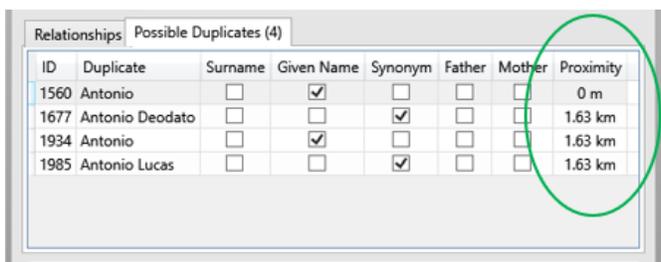
Figure 4.5: Listing of possible duplicates with spatial proximity shown.
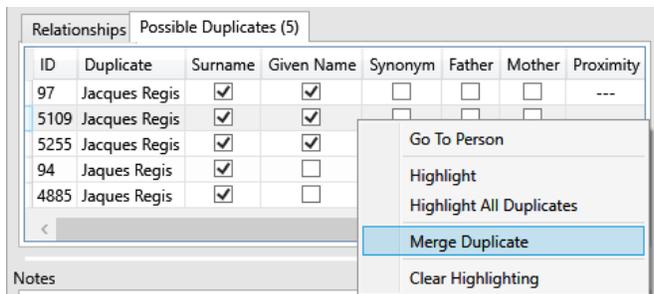
Figure 4.6: Context menu that allows the user to navigate to possible duplicates and to highlight them in the Person Hierarchy.

Linking historical maps to texts

A second identified obstacle to historical analysis concerns the spatialization of family property and the linking of the spatial references in the sources to actual places that can be mapped in a GIS. Achieving this desired outcome was made possible by using information from the notarial records in combination with the toponyms identified on the maps. Because many of the notarial entries record the transfer of property, between named individuals invariably identified by a toponym, it was possible to link the people recorded in the transactions to geographical locations identified on the maps. This process required three phases of work in which the first task was the digitalization, processing and ingestion of the 19th century maps into the system. The second phase was the development of a user interface to support the recording of property transactions and the third was the development of additional query capabilities within the map interface.

All the map processing discussed below was performed using the ESRI ArcMap software platform. This package has a variety of tools that support raster image processing, raster to vector conversion, and map manipulation. The first step towards making the scanned images of the map available for processing was to copy them to a folder on a disk drive accessible to the ArcMap software. Upon ingestion of the digital maps into ArcMap, it was apparent from visual inspection that there were many spurious markings on the scanned images that did not represent actual map

features. These markings included various hand-written labels, notations, and smudges that are artifacts arising from the use and age of the maps that were captured by the scanning process. To make the maps usable within a GIS the initial task was to manually clean the unwanted markings from the images using several raster processing tools within the ArcMap software package.

These 19th century maps were hand-drawn and were created prior to the development of standard coordinate systems. As such, the maps needed to be digitally adjusted so that they matched the actual geography of St. Vincent. This process, called georeferencing, was performed by visually matching and warping points on the map to corresponding points on a coordinate-based modern map or satellite image. These points, known as control points, are used by GIS functions to stretch and warp the original map such that it aligns with the actual coordinate geography (Figure 4.7).

Figure 4.7: Control Point Selection. The points in the northern section were selected so that they minimized the overall distortion of the map.

This georectification process required sufficient control points such that the maps could be modified to the correct projection. In many cases there were features such as buildings and roads clearly visible in the satellite imagery that offered accurate reference points. However, many maps were of areas that are now covered with vegetation and trees that obscure surface features. This can be seen in Figure 4.8 where a scan of a map is overlaid on current satellite imagery. In this example there are many clearly identifiable features that exist today that correspond to those found in the southern section of the map. The land surface in the northwestern section of the map, however, is currently wooded and it is not possible to accurately locate the roads shown on the map in the same way. When situations like this are encountered, control points are added that cause minimal local distortion of the map. As surrounding maps were subsequently georeferenced, the control points for these obscured areas were adjusted to allow the map edges to align as closely as possible. The number of control points per map in our St. Vincent example ranged from 14 to 83 per map sheet.

Figure 4.8: Aerial imagery of modern terrain in the area covered by map 8. Image a) is with georeferencing and image b) is without georeferenced map overlays. Note the close alignment of the current field system with the parcels delineated on the nineteenth-century maps.

The result of the time-consuming georeferencing process was to produce a striking correspondence between the current St. Vincent landscape and the features identified on the

nineteenth-century maps. Many roads, structures, and fields from the maps remain clearly visible in the aerial imagery. In the case of the location of villages this was not too surprising, however field boundaries and traces of the canal system also align well to provide a high level of confidence in the georeferencing results.

Based on the georectified raster images, vector lines were constructed that correspond to individual parcel boundaries and other linear and polygonal features. This raster-to-vector conversion is shown in Figure 4.9. Notice that most of the unwanted markings in the raster images were successfully removed and the boundaries are now clearly delineated. These vectors were then used to create polygons for each of the features. These polygons are the final representation of the spatial extent and location of each feature and parcel that was drawn on the original maps.

Figure 4.9: Detail of map 6 showing (a) the raster version and (b) the resulting vector lines

Each resulting polygon on the geocorrected digital map represents the *geographical information* for a particular feature on each map (Figure 4.10a). In addition to this spatial information, the maps contain identifying numbers for many of the features and visual clues as to what the features represent. On the maps of St. Vincent there are 17 discernable types of feature, four of which, parcels of land, structures, roads, and canals, can be seen in Figure 4.10b. The maps also contain the names of roads, waterways, towns, and some churches that were manually entered into the attribute data. In total, over 25,000 polygons were generated from the 67 hand-drawn maps. Of these, over 21,000 polygons were numbered on the source maps, each requiring manual entry of the number values.
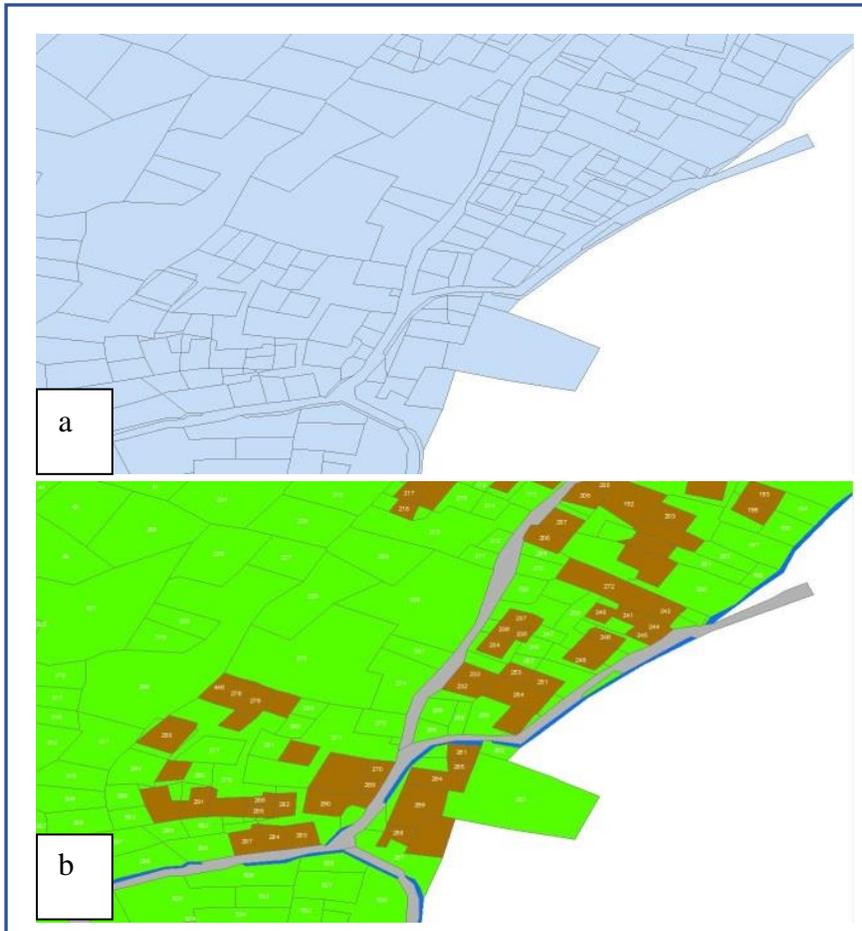
Figure 4.10: Detail of map 6 showing (a) raw polygons and (b) the polygons with parcel numbers color code by feature type. Green polygons identify parcels of land, brown indicate structures, gray indicate roads, and blue indicate canals

The map set comprising 67 scanned maps constituted the entire nineteenth-century survey of the region. Each of the processing steps mentioned above had to be repeated for each separate map resulting in 67 individual maps (or feature classes) within the GIS requiring approximately 180 hours of map manipulation for the set. The resulting maps were then consolidated into a single comprehensive map and mismatched boundaries were adjusted. In addition, features such as roads, that crossed from one map to another were aligned. For most roads, canals, and creeks the alignment process had to be repeated across several maps. With these tasks complete, a comprehensive set of geographical information was available to serve as the backdrop for the events that were recorded in the notarial texts.

The alignment of the map features with the satellite imagery provides considerable confidence that the place names from the map registers match those from the notarial records (Figure 4.11). A subsequent comparison of territory and toponym names confirmed that many of the names on the maps do in fact accurately locate sixteenth-century named references. The alignment also supports anecdotal evidence from current residents of the region who indicated that many current toponyms have existed for hundreds of years (Vester, 2014).

Figure 4.11: Example of the close alignment of mapped structures in brown with those buildings visible in satellite imagery.

For this research, the key to using the maps was that the numbered parcels had to correspond to the map index entries which list the territory (*frazione*) and toponym (*toponimo*) names in which each of the parcels are located. These place names provide the important link that allows the toponyms in the sixteenth-century documents to be spatially located using the 19th century maps. By finding toponyms on the map ledgers that are the same or similar to those in the texts it is possible to know which specific set of parcels may have been involved in a given transaction.

Of course, theory is always more effective than practice. In working with the map ledgers many of the same problems arise as those encountered when working with the sixteenth-century notarial records. The most obvious issue is that the ledgers for nine maps were missing. While there are numbered parcels on these maps there is no corresponding information available for assigning toponyms. The village names shown on these maps themselves offer some implicit information but far less than the level of detail available in the ledgers. Along with missing ledgers there are also ledgers with gaps and with difficult to decipher names, which reduces the confidence that can be placed on matching place names from the sixteenth-century texts with the names from the surveys.

Linking People to Places

Transactions recorded in the notarial records provide the core information that allow links to be established between individuals in St Vincent and the geographically located properties. The user interface used to enter and manage these linkages must therefore address certain information management requirements and must also allow the user to easily specify the details of a given transaction. For the following discussion of the user interface, the document text shown in Figure 4.12 will be used as the example. To facilitate discussion, the details of the transaction have been extracted from the original text and identifying labels have been added. It is also assumed that the records have already been created for the people named in the text.
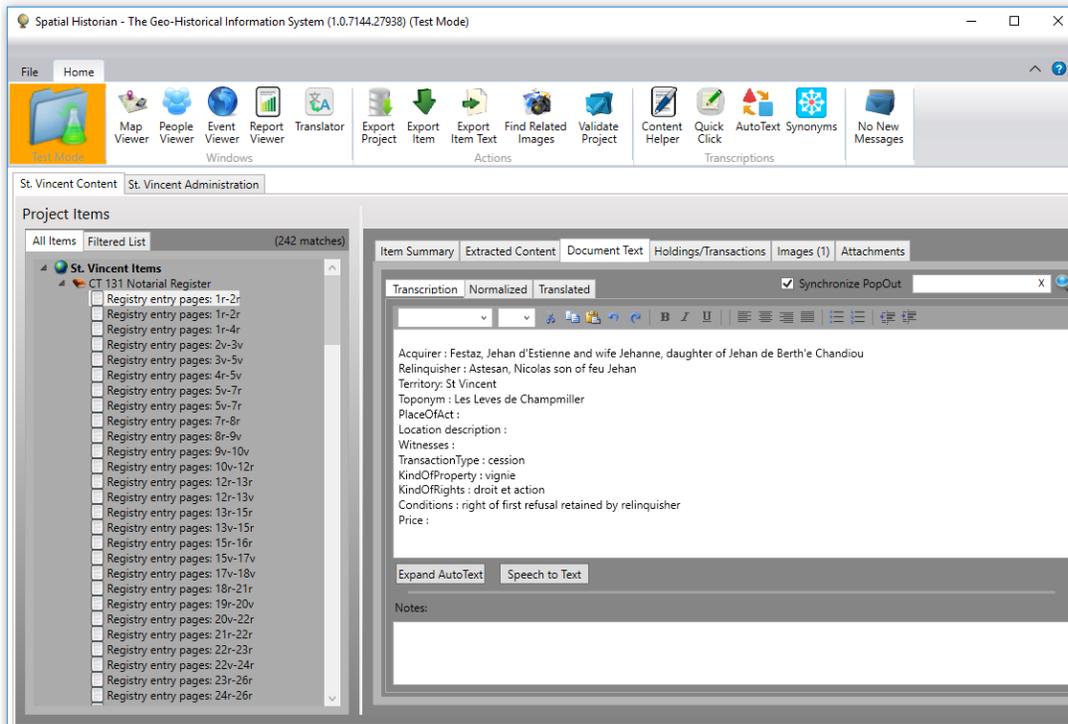
Figure 4.12: Example transaction text. This is a representation of original text extracted from the notarial records accompanied with the specifics of the transaction to simplify the interpretation.

In this example (Figure 4.12) Nicolas, son of the deceased Jehan Astesan, is relinquishing a vineyard, of unspecified size, to Jehan d'Estienne Festaz and Jehan's wife Jehanne. The text indicates that the vineyard is located in an area with the toponym *Les Leves de Champmiller* contained within the broad territory of *St. Vincent*. It can also be discerned from the text that Jehanne is the daughter of *Jehan de Bartholome Chandiou* and that Nicolas' father, who was then deceased, was named *Jehan*. The process of recording these transaction details is done by working with the "*Holdings/Transactions*" tab for the appropriate document section as shown in Figure 4.13. Because this activity requires reading and interpreting the text, it is useful for the user to open the text in its pop-out window as described in Chapter 3.
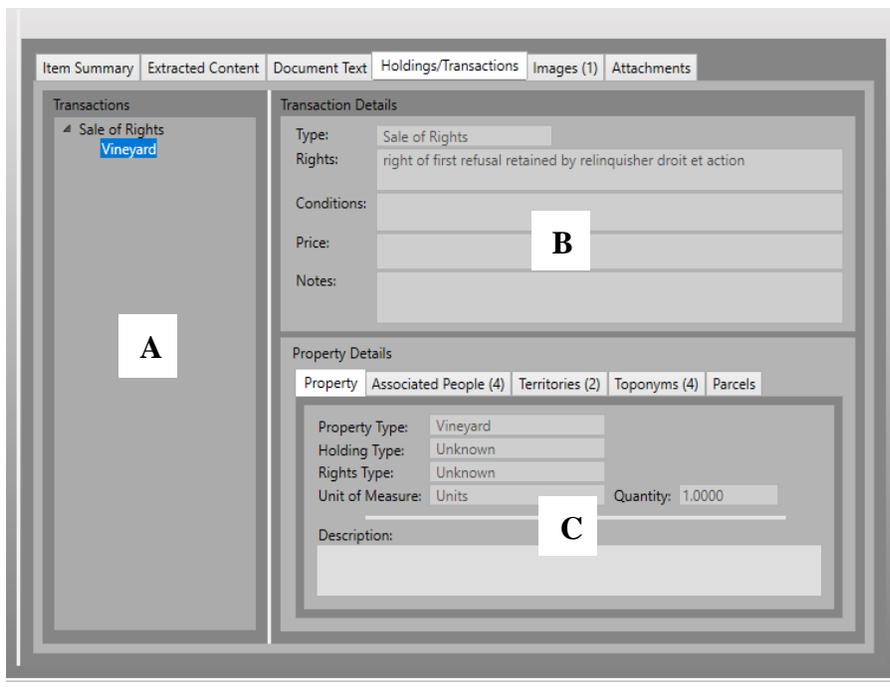
Figure 4.13: Interface used to view and enter property transaction details. A) List of transactions and included properties, B) Transaction details, C) Details about the property involved in the transaction.

The Holdings and Transactions tab (Figure 4.13) is divided into three sections; the left section contains a tree listing the transactions recorded in the text along the properties involved, the top right section contains details about the transactions, and the bottom right section contains details about a specific property involved in a transaction. In this example, the transaction is a cession of rights involving a vineyard. To create a new transaction record, the user selects *Add Transaction* from the context menu. The details are then entered using the pop-up window show in Figure 4.14. The window contains a list of project specific transaction types that the user can chose from. The other fields in the window allow free-form text entry as deemed appropriate by the user. Updates made using the edit window are recorded in the database and are immediately reflected in the upper right portion of the transaction tab.
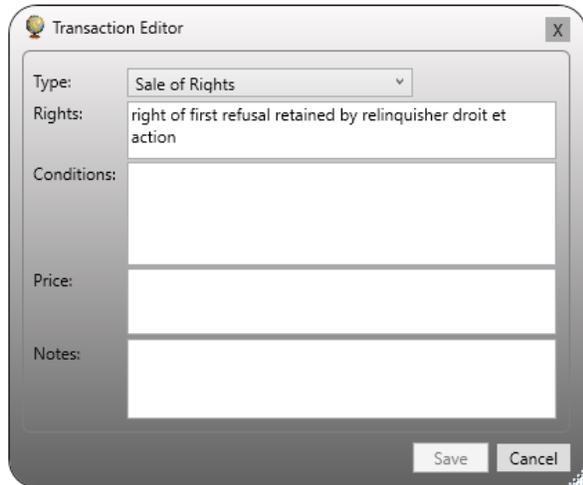
Figure 4.14: Window used to enter and edit transactions.

The next step is to enter information about the individual properties involved in the transaction. The bottom right section of the transaction tab is where the details about the properties being transacted are recorded along with their locations and the linkages to the parties involved. In this example, Nicolas, relinquished a vineyard to Jehan d'Estienne Festaz and his wife Jehanne. To record this information the property is first added by right-clicking on the transaction in the tree list and selecting the *Add Property* option from the context menu. A window then appears that allows the details about the property to be entered (Figure 4.15). This window offers project specific lists of *Property Types*, *Holding Types*, *Rights Types*, and *Units of Measure* to choose from which are configurable by project administrators and system administrators. In this example, the size of the vineyard was not given so a Unit of Measure was set to *Unit* and the quantity was set to 1.
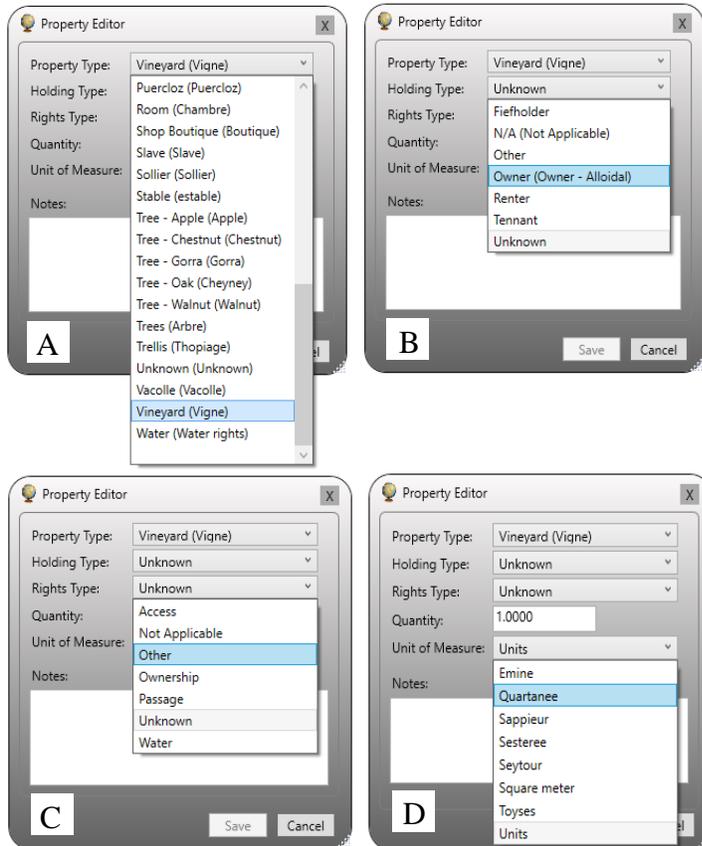
Figure 4.15: The window used to enter details about a property involved in a transaction. The drop-down option selections are configurable per project. A) Property type options B) Holding type options C) Rights type options D) Unit of Measure options.

With the property defined, links to people and places can be added. To add people, the *Associated People* tab is selected as shown in Figure 4.16 and people are added by selecting them from the People Viewer and 'dragging and dropping' them onto the open space in in the *Associate People* tab and a new row for the person is added to the list. Once added the person's role in the transaction must be defined. By default, the role is set to be the *Acquirer* of the property. To change to a different role, the user right-clicks the entry in the list which causes a context menu to appear that presents a choice of project available roles (Figure 4.16). Where possible, the role of an individual within a transaction along with the date and type of the transaction is used by the Spatial Historian to determine the start and end dates of property ownership.
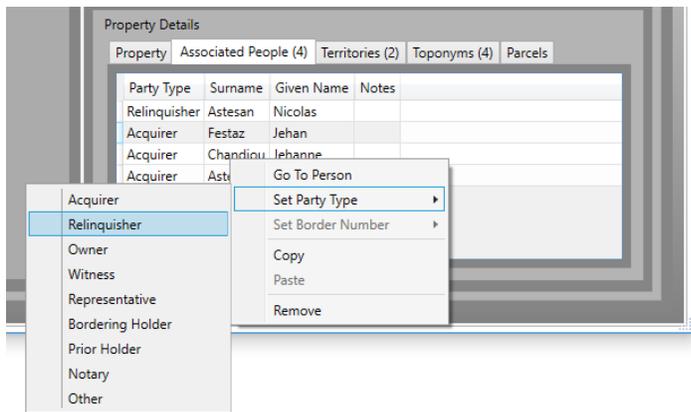
Figure 4.16: Assigning a person's role related to a property within a transaction.

The Spatial Historian allows a researcher to enter duplicate names for individuals and these duplicate entries can later be merged as additional defining information is gathered. Thus, in the case of property transactions, the researcher can link a property to a specific individual or to multiple individuals. In cases where duplicate individuals are associated with the property and the duplicates are later found to be the same person, the *Merge* process can be performed from the *People* window. When the merge is performed, the specified property is linked to the resulting person. In cases where it is found that the possible duplicates are not the same, the link to an incorrect person can simply be removed by selecting the *Remove* option from the context menu as shown in Figure 16.

In addition to recording who was involved in the transaction of a property, the Spatial Historian allows a researcher to identify where the property was located (Figure 4.17). It is important to note that while the headings for the property locations in this example are located on tabs labeled *Territories*, *Toponyms*, and *Parcels*, the spatial hierarchy is customizable per project. For example, some other project may choose to define a four-level hierarchy based on *Province*, *City*, *Parish*, and *District,* in which case the tabs adjust automatically to the alternate specification.

Figure 4.17: Specifying the location of a property involved in a transaction

The procedure for adding a location identifier is similar to that used to associate people with property transactions. To add a location the researcher opens the *Map Viewer* and expands the map feature list to find and select the appropriate item. The map feature is then dragged and dropped onto the grid. The system automatically validates that the appropriate type of map feature, which in this example is *Territories,* can only be dropped on the *Territories* list and only *Toponyms* on the *Toponym* list.

If any ambiguity in the place names associated with the St. Vincent project is found, the system allows multiple places to be specified and a level of confidence to be assigned individually to each entry. The choice of confidence level is based on the individual researcher's interpretation of the text and the integration of other contextual information that may be available. The confidence level is set to *Medium* by default but can be changed by right-clicking an entry in the list and selecting an appropriate setting (Figure 4.18). The confidence level can be used for subsequent reporting and as a filter for spatial queries in the *Map Viewer*.
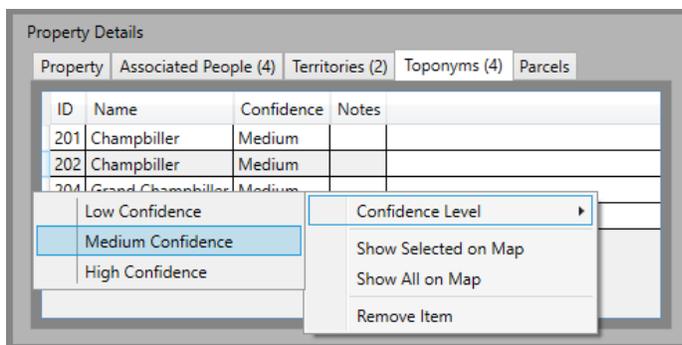
Figure 4.18: Changing the confidence level for a property location.

Spatial Queries - Spatializing Kinship Groups

Information collected using the custom features of the Spatial Historian can be used for a variety of purposes including the examination, for example, of how often different types of property in differing areas change ownership or are identified for various land use. The system necessarily supports the goal of the St. Vincent project to "spatialize kinship relations" in the region by allowing the researcher to examine where and when members of different family groups purchased and sold properties. The tight coupling of the documents and the people and places they reference, allows system users to drive the exploration of the spatial content from the documents, the people and families, or from the maps. The map-driven spatial query functions are performed by selecting the *Query* tab on the left side of the Map Viewer (Figure 4.19).
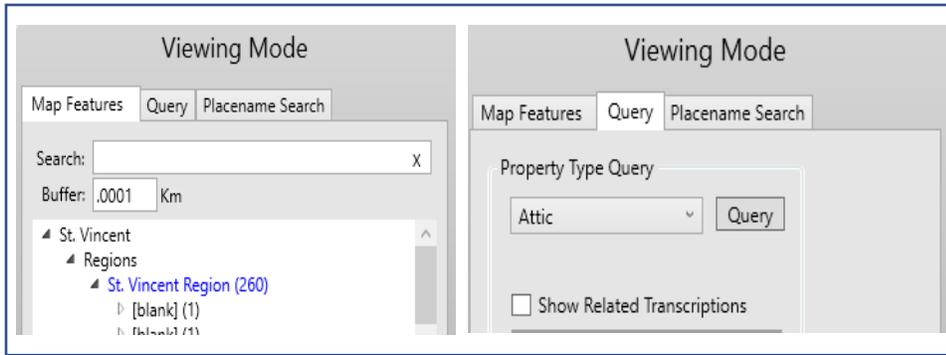
Figure 4.19: Accessing the query functions on the Map Viewer is accomplished by switching from the 'Map Features' tab to the 'Query' tab and from there initiating either a query based on a user-defined polygon or based on specifying a property type.

Data exploration initiated from the map interface can be driven by specifying bounding polygons on the map interface to initiate spatial queries or by selecting a search based on property type or land-use. Figures 4.19 and 4.20 demonstrate an example of a query based on a user defined polygon. In the example, the user has requested details about the features contained within the yellow polygon. Upon running the query, the parcels of land are plotted, and the resulting *Spatial Query Details* pop-up window is presented with the listings of all people and containing map features. From the listing of people, the user can navigate directly to corresponding records in the *People* window. Similarly, from the listing of Territories and Toponyms, the user can navigate to corresponding entries in the map feature hierarchy of the *Map Viewer*.
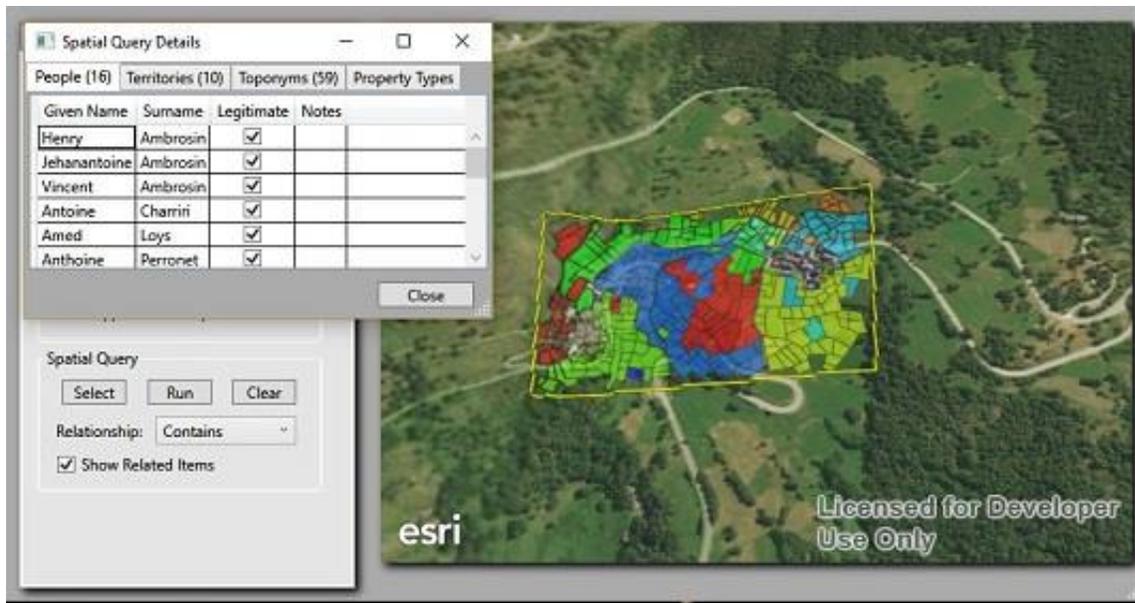
Figure 4.20: A spatial query listing the people and place names found within a bounding polygon

Performing a query based on property type causes the system to construct parcel polygons for all parcels that are linked to the specified type (Figure 4.21). Optionally the user can indicate the minimum level of confidence to be used when querying. For example, when the user specifies that the minimum confidence level is *Medium*, property linkages recorded with a *Low* confidence level will not be included in the results. As with the spatial query results, the user has the option to have the parcels and related people displayed in the Spatial Query Results pop-up window to facilitate additional data exploration.
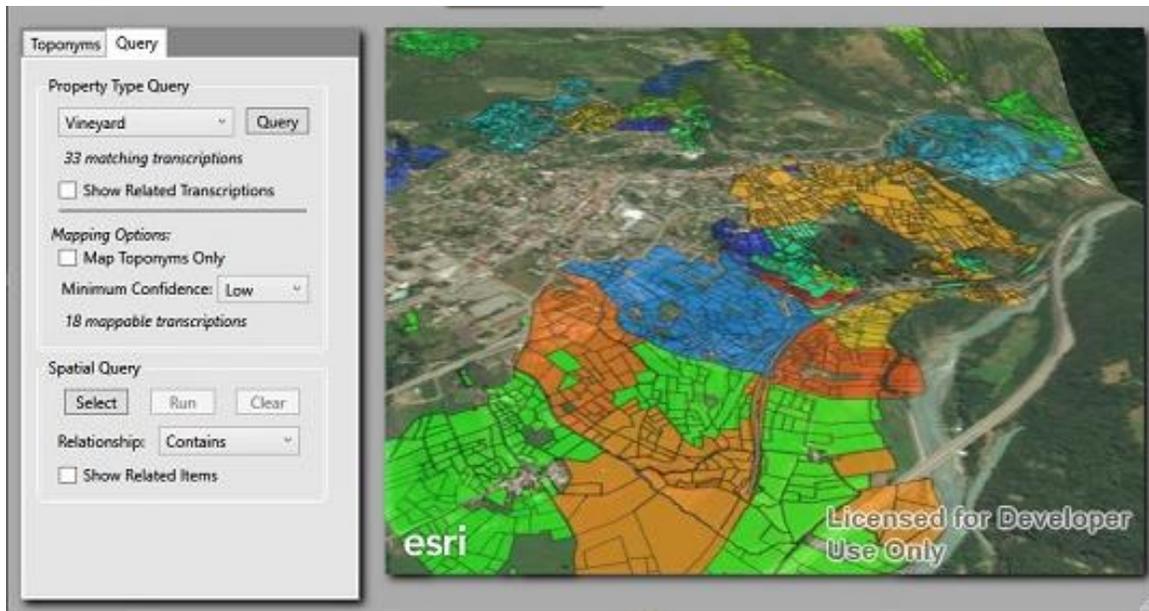
Figure 4.21: Land use query result listing parcels linked to grape production.

Finally, spatial queries based on selected people are performed using the People Viewer window. Queries are initiated by right-clicking on an entry in the hierarchy of people to open the context menu. From this menu, the user has the option to map the holdings for the selected person or for different relationships groups (Figure 4.22). The resulting spatial queries based on the selection of a person are similar to those for map driven queries; related parcels are rendered, and the Spatial Query Details window is presented with tabular listings of matching people and places.
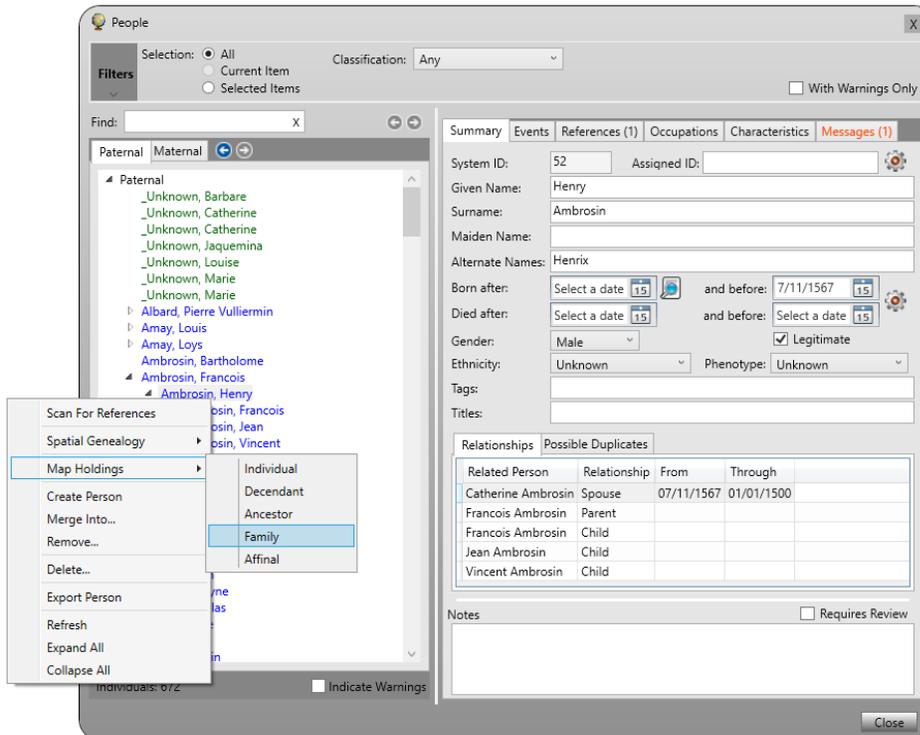
Figure 4.22: Using the People Viewer to map landholdings for individuals or kin groups

Case Study Outcomes

For the St. Vincent case study, the system performed well in spatializing kinship relations within a rural mountain community during the early modern period. The system was capable of comparing complex family dynamics within, and between, rural families and within a variety of social settings (Appendix B). The Spatial Historian in this instance provided both the spatial and text data handling capability required to aid transcription, text extraction, and text recording. Furthermore, one obstacle impeding the desired research goals for the St. Vincent project was "the difficulty in determining whether individuals identified in various records who have the same name (and perhaps the same father's name) are in fact the same person, or whether they are different people" (Vester, pers comm.). The Spatial Historian was able to populate a names list with associated dates and similar name spellings against which other recorded names could be

compared to help facilitate name disambiguation. A second obstacle noted by Vester concerned the spatialization of family property and the linking the spatial references in the sources to places that can be mapped in a GIS (Appendix B). In this instance the Spatial Historian was enhanced with the functionality to link property ownership as recorded in the notarial records with individuals, and by extension, to kinship groups. These successful technical results demonstrate the ability of the system to record and manage property transactions and to link them to geographical locations.

CHAPTER 5 -Applying the System: The Slave Societies Digital Archive

The second case study examines the work of the Slave Societies Digital Archive (SSDA), which is dedicated to preserving ecclesiastical and secular documents related to Africans and those of African descent in the slave societies of the Americas. The project, directed by Dr Jane Landers at Vanderbilt University, was initiated in 2003 (Vanderbilt University College of Arts and Science, Department of History, 2018). The primary directive of the Archive is to "preserve and freely disseminate invaluable documents for the history of Africans and their descendants in the Atlantic World" (Vanderbilt University, 2018). As such the initial work of the SSDA focused heavily on capturing digital images of the actual endangered historical documents in order to preserve their content and to make them available for subsequent research. The documentary materials in this archive are comprised primarily of the ecclesiastical records of baptisms, marriages, and deaths. The archive currently contains over 310,000 document images from archives in Brazil, Colombia, Cuba, and the United States, with references to an estimated four to six million individuals. Funding for the SSDA has been provided primarily by grants from National Endowment for the Humanities, the American Council of Learned Societies, the Andrew W. Mellon Foundation, and the British Library's Endangered Archives Programme.

The SSDA is a partner in the ongoing Mellon Scholarly Communications grant entitled *Enslaved: People of the Historic Slave Trade*, which links the project to other ongoing initiatives at Michigan State University, Harvard University, Canada's York University, the University of Colorado, and University College London in a collaboration to create a Linked Open Data hub.

This is collaborative seeks to reconstruct a more complete picture of the Atlantic Slave Trade and requires that the data collected by the various research partners be consolidated into a globally accessible information network that can be used to support studies that expand beyond the bounds of the individual data collection projects.

Applicability as a Case Study

Several characteristics of the SSDA project stand distinct from the St. Vincent project as a case study. First, the SSDA project involves a very large number of documents and managing the extracted content from the documents challenges the scalability of the Spatial Historian in terms of the storage requirements and the performance of the user interface. The exceptionally larger SSDA collection necessitated three primary system upgrades over and above that developed for the St Vincent study. These system changes included the development of a more streamlined record entry user interface named the *Content Helper; e*nhanced support for differing historical document types; and internal system performance upgrades involving a shift from standard ESRI mapping services to a customized spatial data management and display components. Second, the ecclesiastical documents in the SSDA collection provide only limited snapshot of information about the individuals recorded as the vast majority of individuals are only referenced once in the corpus, for example at the time of baptism or marriage. For this reason, aggregate reports and the development of unique visualization techniques were required to support data analysis of patterns of activity over time and space. Third, the research goals of the SSDA introduced broad historical events into the analysis, called *Shared Events*, and the interaction between these events and the effects on individuals became an important part of the data processing requiring a new set of analytical techniques. These relationships linking shared events and the individuals recorded are particularly visible in the shared experience of a slave voyage, for example, and its impact on the

enslaved, the crew, the ship owners, and the slave owners. Fourth, the involvement of the SSDA project in the *Enslaved: People of the Historic Slave Trade* initiative required the additional development of system functionality to enhance the data integration and collaboration with the other external project partners. This collaboration required the development of an extensible data export functionality. Finally, the large numbers of individuals referenced in the SSDA collection necessitated the creation of functionality to support Social Network analysis based on the familial and social relationships recorded in the texts.

The Primary Source Materials

The record of Mateo's baptism mentioned at the beginning of this dissertation is typical of many records to be found in the SSDA volumes in that it contains additional information beyond the names of participants in the ceremony (Figure 5.1). For example, it can be inferred that the elder Mateo and Ysabel were partners, though it is not known if they were married or whether their marriage would have been recognized by the church. Also, it is implicit that all of those mentioned in the entry were born prior to January 20th of 1590 and (likely) died after that date. It can also be inferred that Alonso Çuares de Toledo and Baltasar Surdebiera were slaveholders, with Çuares de Toledo owning at least three slaves including the newly baptized Mateo, and Surdebiera owning at least one. Juan Ponce Roman can be assumed to have been a member of the clergy, or at a minimum had an affiliation with the church where the ceremony was held. Francisco de Bihera is little more than a name, although it is plausible to assume that he was a freeman since no mention is made of his owner. Because *Sagrada Catedral de San Christobal de La Habana* still exists today, it is also possible to place the individuals mentioned (with the possible exceptions of the slaveholders) at a specific location on a specific date.
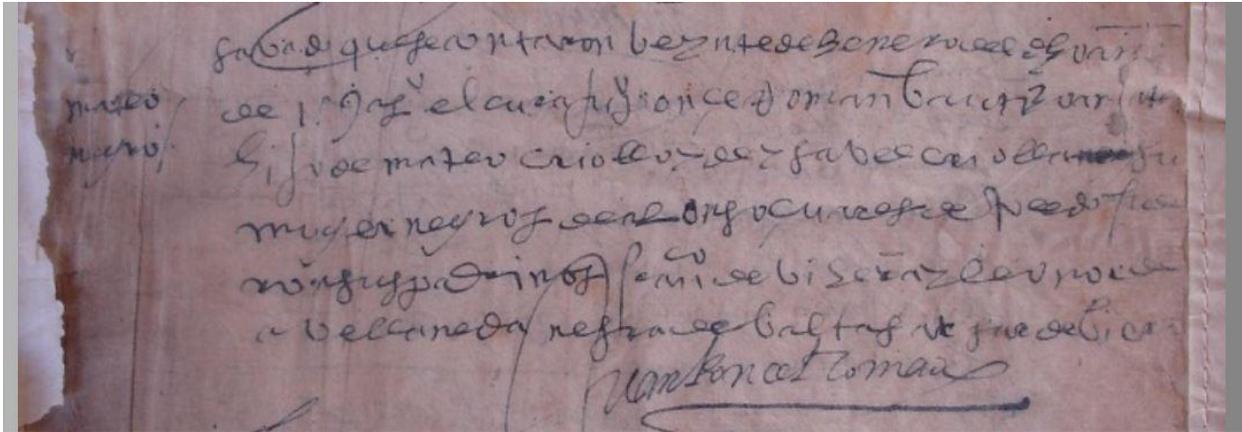
Figure 5.1: Baptism entry for Mateo Criollo - January 20th, 1590. From Libro de Barajas: Bautismos 1590-1600

Taken in isolation, the six lines penned on January 20th, 1590 offer valuable information, but they are not enough on their own to develop a deeper understanding of the individuals, let alone a larger narrative of society of the time. The true worth of this snapshot lies in assembling the contents of other materials and linking the record information to large continuous and longitudinal sets of baptismal, marriage, notarial, and other administrative records. In this way multiple references to people, places, and events over time can be linked and cross referenced. As an example, other entries in the archives of *Sagrada Catedral de San Christobal de La Habana* referring to Juan Ponce Roman might be used to frame his tenure. Likewise, additional references to Alonzo Çuares de Toledo and Baltasar Surdebiera may offer supplementary information about the extent of their slaveholdings. Most importantly, examining the corpus of data in aggregate may provide a broader perspective of the structure and extent of slaveholding in the region at the time and proffer a glimpse into the lives and extent of the slave families themselves. The Spatial Historian provides the means to ingest and manage these records to support extensive additional analysis.

The use of the Vanderbilt SSDA project as a case study was to further challenge the capabilities of the Spatial Historian system beyond the work of the St Vincent project as the SSDA

project quickly evolved into the application of the Spatial Historian as an integral component of their major research program. This evolution demanded that the software be transformed quite rapidly from a 'theoretical' research prototype into a fully functioning production system. One outcome of this transformation was the necessity to shore up existing functionality and to develop new, and previously unanticipated capabilities. Some of the new features were driven by the demands of the project, while others grew from knowledge gained from the use of the system itself. In other instances the goals of the project acted as an inspiration for envisioning new ways of extracting knowledge from the data.

The materials in the SSDA archive then created several additional challenges to the system due to their deteriorating physical condition, their paleographic style, and the sheer number of individuals they reference. The condition and paleography of the documents make automated OCR untenable requiring the project team at Vanderbilt to manually transcribe the content from the images. The SSDA project currently has two volumes transcribed and it is these that form the basis of this study. The Vanderbilt University team are now at a point where the group need to extract the specific details about individuals referenced in the documents, along with the places, events, and relationships that made up their lives. The Spatial Historian was used to support the extraction and management of this information and was employed in the actual analysis to develop a spatial knowledge about the slave trade based on the implicit spatial content of the documents. Once extracted, the system was used to explore the social and familial relationships of the recorded individuals and developed an understanding of the events and spatial linkages that brought them together. Thus, references to slaveholders have been used to determine if patterns of land holding sizes can be ascertained. In addition, ethnonyms and occupational titles have been examined using the system to develop an understanding of the movement of people affected by the slave trade

including slaves, slaveholders, and the supporting populations of clergy, merchants, and laborers. The interactive mapping capabilities of the system support these various types of analysis through the implementation of interactive spatial analysis features such as visualizing spatial changes in the life events of slaves over time. The interactive map interface also supports embedded hyperlinks to related materials and 3D models are supported as attachments to document transcriptions as well as to map features.

Further Applications of the Spatial Historian

The demands of the SSDA project led to several conceptual and technical changes to the system and in particular prompted the need to record and incorporate *Events* into the system. These events include both life events for individuals and events shared by many people. The high volume of records that had to be processed by the SSDA project led to the need for greater data validation to be integrated into the data entry process along with the development of a user interface, called the *Content Helper*, that supports more rapid entry of large numbers of records. In addition, the data content entry was further automated to include the start and end dates of relationships and the automatic creation of relationships based on the extracted content. The variety of languages used in the source materials also led to enhancements to the synonym processing function that included a reconfiguration of the spelling variation and diacritics capability. Finally, the relationship between the SSDA project and the MSU *Enslaved* project led to the development of an extensible data export capability. Several of the most notable enhancements, including the Content Helper and event visualization are discussed in greater detail below. Videos of Content Helper, the Event Viewer, and other enhancements can also be viewed online by following the links provided in Appendix A.

The Content Helper

Because of the very high volume of document images in the SSDA collection, creating records for people along with their associated relationships and references became problematic. The time required to process even a single volume of this large collection was prohibitive in time and resources. Due to the need to increase throughput a customizable window called the Content Helper was created in the Spatial Historian. Rather than requiring the user to enter details for each person individually and to then manually create event and relationship entries, this window provided a means for preforming all of these tasks in a single automated process. Importantly, the configuration of the Content Helper adjusts dynamically to optimize the work-flow based on the type of document being processed. It accomplishes this by modifying the number of people prompted for, and the roles that each might have played based on the type of event recorded in the document section currently being processed. For example, Figure 5.2 shows the Content Helper as it is displayed when processing a baptism record. In this instance, the window has an input 'block' for the person being baptized, labelled 'Self', along with blocks for up to two parents, two godparents, three slave holders, the official presiding over the service, and a scribe. The bottom portion of the window allows the user to specify additional details about the event including its location and the date range in which it occurred. Adding individuals to the Content Helper can be accomplished in several ways: the names can be typed in directly, names can be dragged and dropped directly from the transcription text, existing people can be dragged and dropped from the People window (discussed below), or they can be added using the *Quick Click* feature.

Figure 5.2: Content Helper configured to extract content from a baptism record.

As the names of individuals are added to the Content Helper, the system makes an initial determination of gender and simultaneously adjusts the display to indicate whether linkages are expected. In a baptism example, when a person is entered in the *Owner* block, the '*Owner'* label background color changes to green to indicate that one of the other individuals must be identified as a corresponding slave. For baptism entries this could be the person being baptized and/or any of the parents or godparents. Figure 5.3 shows an example with *Owner/Slave* relationships established between each of the Godparents and the *Owner* named Hernando Manrique. The background color on the role title of the Godparents indicates that the corresponding god-children must still be identified. Establishing these linkages is accomplished by dragging the related person block onto the block of the person expecting a relationship. Note that as the user moves the mouse

over a block, the border of the block turns blue. At the same time the border of the blocks of any related people turns orange. Therefore, as represented in Figure 5.3, the user mouse is over the block for Hernando Manrique and the borders of the related blocks for his slaves are orange.

Figure 5.3: Content Helper with Owner/Slave relationships established between the Godparents and the Owner named Alonso Çuares de Toledo. The green background on the role title of the Godparents indicates that the corresponding God-Children must still be identified.

Once all the people referenced in the document are assigned to their corresponding roles and the additional location and date details are entered, the user can click the *Process* button to initiate the creation of the record for the event, along with the records for the people, their relationships, and their attendance at the event.

Event Visualization - Events, Stacks, Columns, and Clouds

Whereas the St. Vincent case study focused on understanding family structure and property holdings, the SSDA project is concerned more with individuals and their significant life events. Building on Ethington's (2007, p. 483) and Hagerstrand's (1970) work on space-time paths (Figure 5.4) enhanced visualization has been added to the Spatial Historian to handle specific SSDA analytical questions associated with the project. Whereas Hagerstand's space-time cube as it is normally constructed, represents a sequence of movement of an individual entity over time and space it is not capable of representing the coincident movement of groups of people nor the impact (expressed in terms of the number of people) that such movement had on surrounding peoples and places. Rather than a visualization that focuses primarily on individuals the visualization developed for the SSHA project is location and event-type focused. For historical research, the broader interpretation of *what* was happening and *where* can provide a more encompassing view of activities and movements over space and time. Recognition of this project interest led to the development of a novel event mapping interface within the Spatial Historian that uses the vertical axis of the 3D map interface to represent time and a dynamically-created event symbology to indicate event duration and the number of individuals involved. This approach to the mapping of events differs from the mapping of locations in that in addition to spatial coordinates, event mapping contains the temporal extents as well.
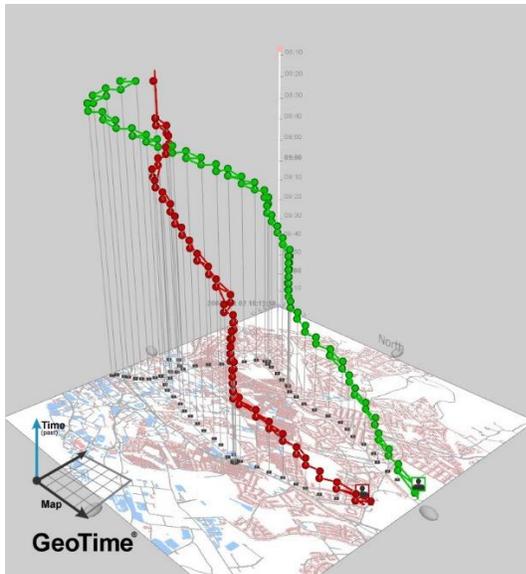
Figure 5.4: Space-time cube example.
(Source http://www.geog.leeds.ac.uk/courses/level3/geog3150/lectures/lecture9/lecture9.html)

The basic component of this visualization approach is the representation of the Event. In this model, beyond the typical space-time representation, events are assigned two additional dimensions, summed up as their impact. For this discussion, impact is defined as the number of individuals directly affected by the event and the immediate duration of that effect. As shown in Figure 5.5, the representation of events takes on the form of lozenges.
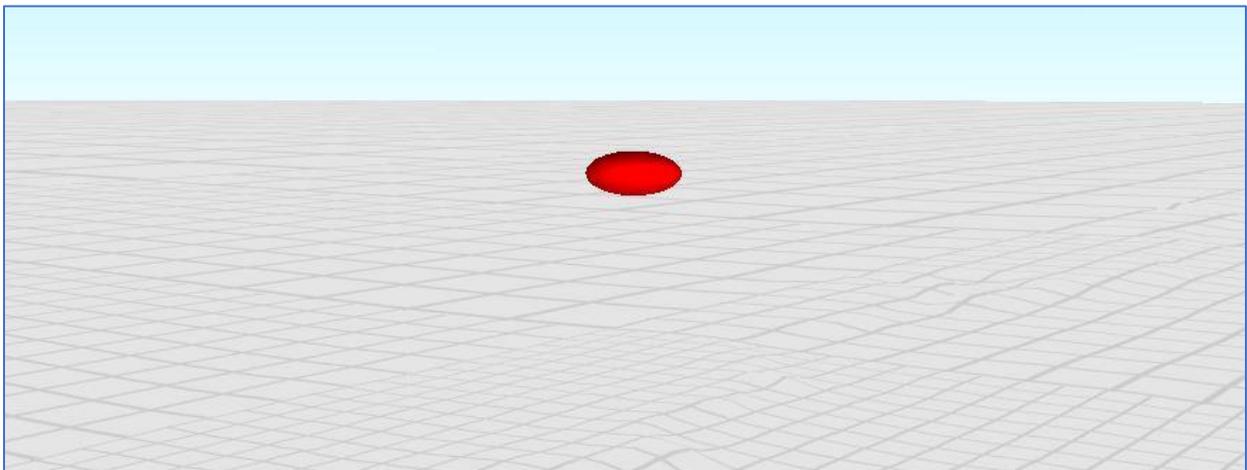
Figure 5.5: Representation of an event hovering above a 2-dimensional plane. The impact if the event is represented by the height and width of the symbol.

The *Event Stack* is the representation of a series of events – with some common thread – occurring over time at the same location (Figure, 5.6). For example, an event stack may be comprised of a series of baptisms occurring at the same church. Another example may be the recording of property transactions occurring at a notary's office over a span of time. The sequences of events occurring at the same location provides common threads of emotional or cultural connectivity between individual participants who may have never actually met. The temporal density of the events an individual stack contains and its duration will help define its social impact and tenure.

Applying Ethington's description of place as "… collective phenomena, transformed by the sentient bodies that inhabit, know, or recognize them. Places are the condition of possibility for human culture itself: 'To be cultural, to have a culture, is to inhabit a place sufficiently intensively to cultivate it .... Culture is carried into places by bodies." (Ethington, 2007, p. 482), the *Stack* is then, the marker of "place" – the location where culture is created. Some locations may have people drawn to them because of their ascribed importance or imposed importance such as a church or a border checkpoint. Other locations may be important because people go there of their own volition; for example, a pub or a grass field that has been adopted for play. Still other locations may be important because of their natural advantages such as a harbor or a river confluence.
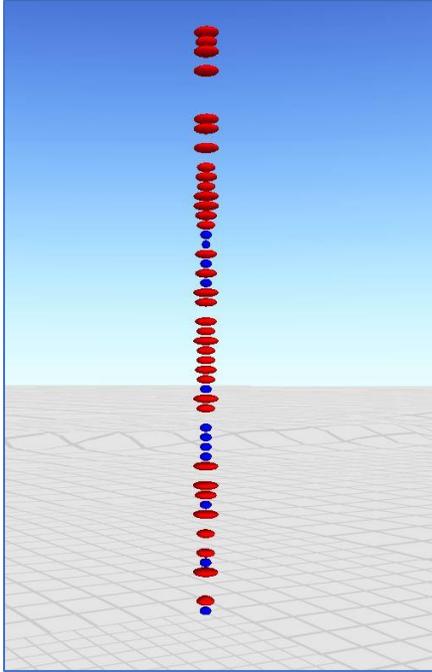
Figure 5.6: An event stack hovering over a 2-dimentional plane. The red and blue symbols represent different event types occurring at the same location.

The *Event Column* is the area of *Convergence,* an area of impact that is affected by the gathering of people over time as shown in Figure 5.7. The flow of individuals acting out common cultural roles imbues the area with a significance. The grounds of a church and the surrounding walkways may become informal forums for social interaction and places where, for example, neighbors chat or merchants discuss the current state of trade. A place and area of place making, community, and memory may develop and be shared by those whose lives intersect at a location albeit at different times. This intersection is unknown to the individuals, who may have never met or who have a common experience of place. These are the footprints of places extended over time. Events in the *stack* affect the surrounding community and the sense of community that develops in its environs may in turn affect the stack by drawing more activity to it.
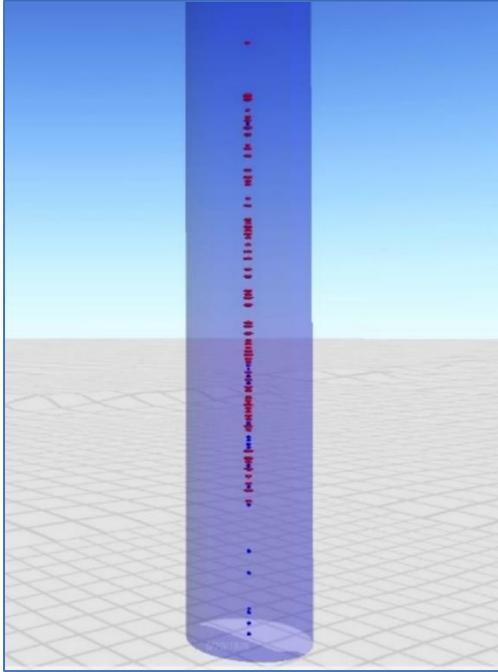
Figure 5.7: The event column showing the area of convergence of individuals.

Event Stacks and Columns are not isolated in their spatial context and by visualizing large numbers of events occurring across multiple Stacks, a *Cloud* of events can begin to emerge – one not dissimilar to the 3D point clouds produced using technology such as lidar or photogrammetry. In medicine, 3D Models constructed from CT scan slices are used to visualize the internal structure of organs for teaching and diagnostics. Similarly, point clouds have been used to analyze dinosaur skeletons from fossilized bone. The internal structure, therefore, of point clouds can provides insight that may be unavailable when only looking at a surface topography. Within the context of historical research, point clouds may also be applied as a means of visualizing the structure and the motion of events over time across a given region. For historical research this presupposes a volume of event data that supports the creation of Event Clouds of sufficient density that patterns ca be discerned. As will be seen in the case studies used for this research, it is the opposite issue that can the problem; rather than having enough events (points) to develop patterns, the problem

in practice becomes one of extracting sufficient events from the documents to develop anything resembling a "cloud" at all.

Events, *Stacks*, and even *Clouds* do not occur in isolation; they are often (if not always) influenced by external events. External factors could be naturally occurring events such as a volcano or earthquake, or human mediated events such as the inadvertent transport of a plague, or intentionally initiated events such as a war. For the purposes of this discussion, these external events can be characterized in several ways. The first is that these types of events affect a significant number of people who may otherwise have no common connection, acting as a links between individuals and communities and creating shared experiences. A second characteristic is that these types of events have a definable temporal extent in that they have a start time and an end time. The duration may be very short, as in the case of an earthquake, or may last for months as in the case of a plague with the impact of the event varying throughout its existence. A third characteristic is that these events and their resulting effects have a definable spatial extent. This is the area within which individuals are impacted or influenced in a common way though the actual impact on individuals may vary widely based on factors such as social status, wealth, access to resources.

Because these types of events create common linkages between individuals who may have little or nothing else in common, they are referred to as *Shared Events* which cross-cut societal structures and the *Stacks*, *Columns*, and *Clouds* of otherwise un-related places. These events may disrupt and rearrange the norm, or they may connect individuals but in either case they provide shared experiences and points of departure for the future. To borrow a term from astrophysics, the planes where the upper and lower temporal boundaries of shared events intersect the events stack (or cloud) is referred to as the *Event Horizon*. The beginning horizon would likely be a distinct

point in time while the events following the ending horizon may continue to show the effects of the shared event as the region gradually returns to its original or new status-quo.

Figure 5.8 represents a shared event cutting across an events stack, clearly showing the timeframe and spatial extent of its impact. The inset shows a subset of events in the stack that would have been immediately affected by the event given their time and position and location of occurrence.
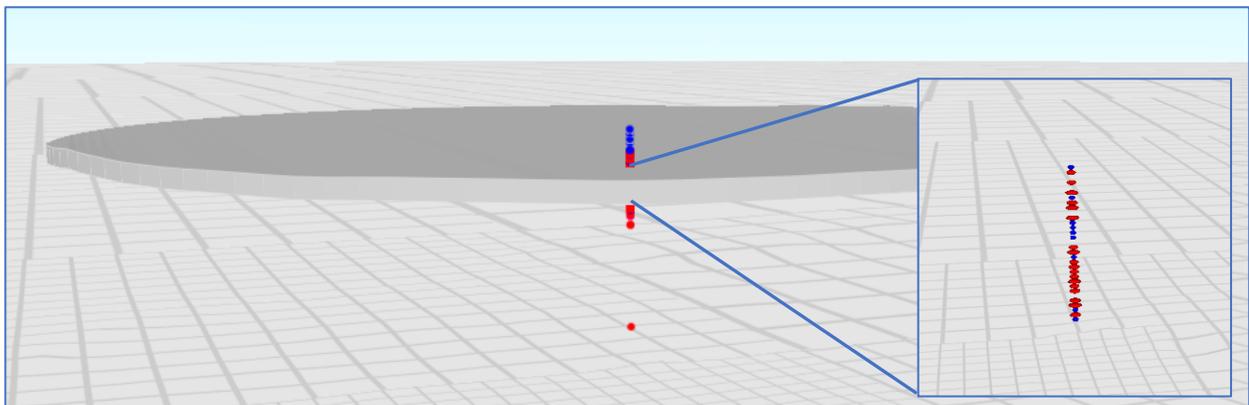
Figure 5.8: A hypothetical Shared Event cross-cutting an Event Stack. Inset shows the subset of Events immediately affected by the Shared Event.

The figures presented above show events conveniently occurring over a relatively flat terrain and thereby allow the vertical or time component to have a consistent base for representing all events in a field of view. That is to say, if the time value at the surface is set to the year 1600 AD, an event represented on the surface will have a time component of 1600 AD. However, consider the representation of a set of Event Stacks located in a mountainous region, such as the one shown in Figure 5.9. In this setting, where the locations of the stacks and therefore the bases, are at different elevations, the vertical positions for 1600 AD events would be different. This affect would cause a comparison of the timing of events across the viewing area to be difficult for a

consistent horizontal plane is required for the vertical displacement of events to be consistently visualized.

Figure 5.9: Simultaneous events positioned on the terrain

To accommodate this situation the system allows the user to turn off the elevation information for the map, and thereby convert the scene from a three-dimensional representation to a flat representation as shown in Figure 5.10. Rendering the viewing area as a flat surface allows the Event Stacks, Columns, and Clouds to share a common vertical base and a consistent representation of time over the field of view.
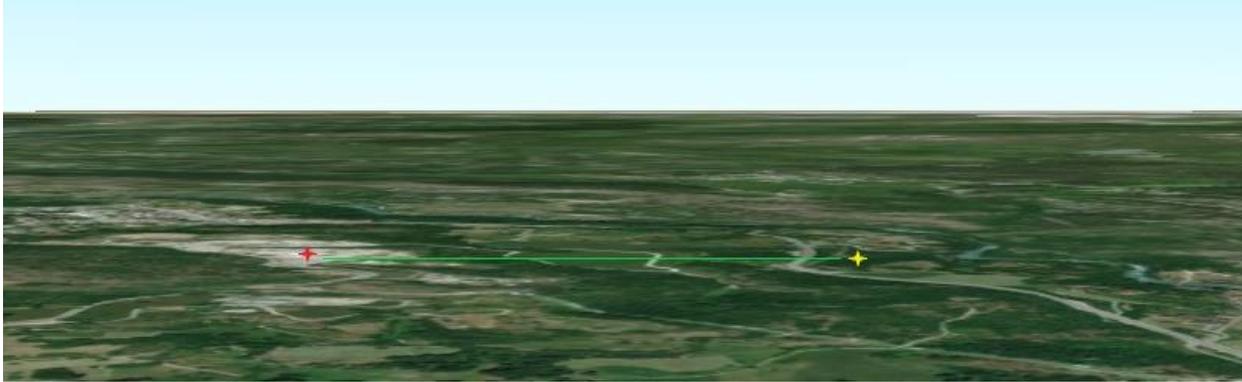
Figure 5.10: Visualization of simultaneous events on a flat surface

There are, however, cases where the vertical displacement caused by such variation in surface elevation may be beneficial for visual analysis. Take for example a set of Event Stacks at different locations and at different base elevations within a viewing area. If some of the stacks are made up of dense sequences of events while others are sparse, any possible variation in terrain as a conditioning factor will not be apparent. By adding the terrain, some stacks may be seen to be located on higher or steeper terrain where others are located on flat terrain or at lower elevation. This observation may lead to the formation of questions about population density or suitability for agricultural production and how these factors may have affected event occurrence. In another example, it may be observed that event stacks at some locations exhibit gaps in their sequence while those at other locations do not contain corresponding gaps. By viewing the stacks in the context of their elevation may indicate that a shared event may have affected the viewing area yet that its affect was elevation or terrain specific.

By applying these techniques to visualizing events contained within a baptism register included in the SSDA case study, each baptism is spatially located at the coordinates of the church and temporally on the vertical axis of the map at a scaled height based on a user set date range specification as shown in Figure 5.11. Likewise, the height of each individual event marker is scaled to the duration of the event in days and the width of each marker is scaled to the number of

people attending the event. The system allows the user to produce Event maps for selected locations by person or family hierarchy, and through the manual selection of individual events.
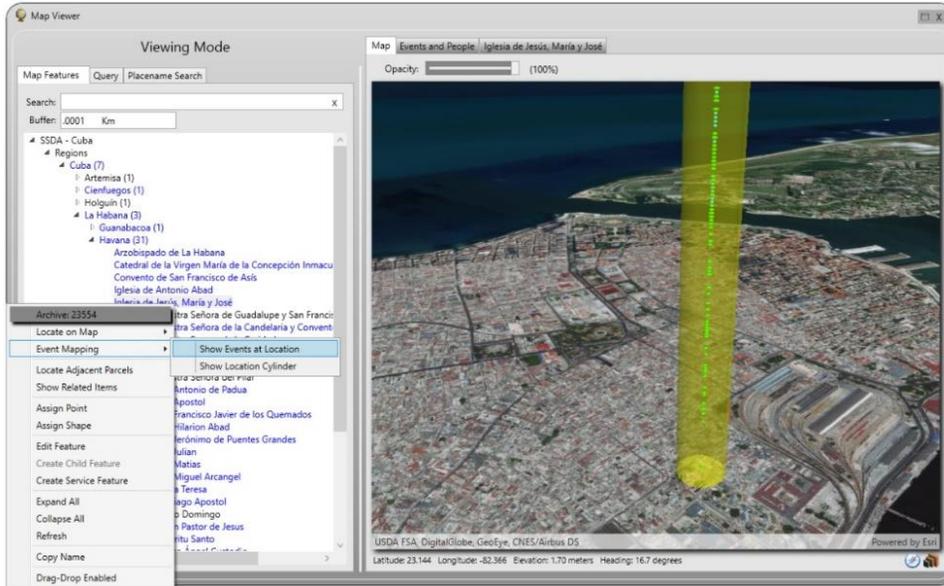
Figure 5.11: Example of the mapping of baptisms that occurred at Iglesia de Jesús, María y José between 6/20/1828 and 12/5/1828.

Shared Events can also be mapped along with the life-events of individuals to explore their possible interactions. In Figure 5.12 a shared event in the form of the arrival of a slave ship, represented by the gray plane, is shown along with the stack of baptism events recorded at *Iglesia de Jesús, María y José*. From this interface the user has the option to filter the baptism events to show only those that coincide with the ship's arrival plus or minus a user specified temporal buffer. This functionality can be used to provide a visual identification of the interaction of events.
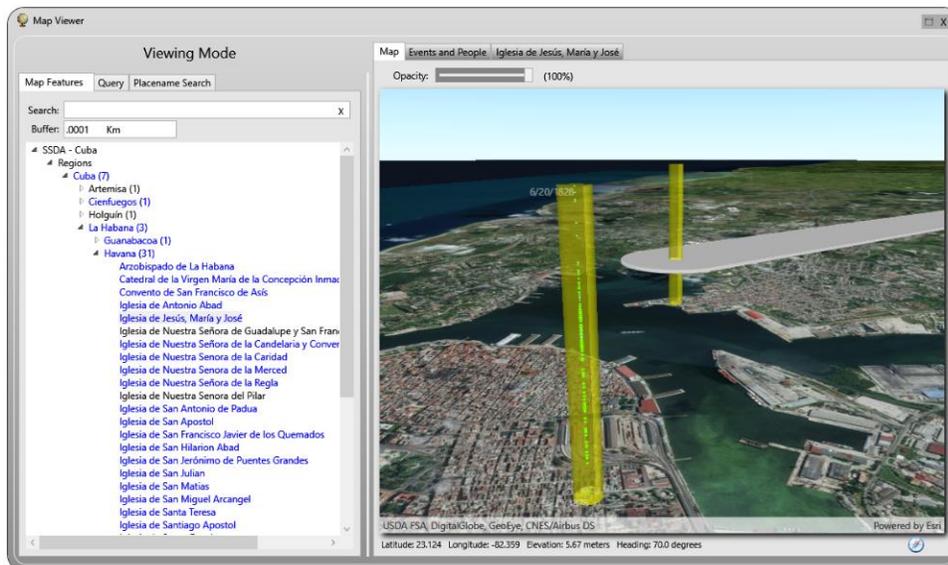
Figure 5.12: A Shared Event in the form of the arrival of a slave ship represented by the gray feature display along with baptism events from Iglesia de Jesús, María y José.

Document Filtering

The large volume of documents in the SSDA collection and the need to manage the corresponding large volume of individuals and events, required the development of new techniques that allowed the system users to filter the contents shown on the screen to only what is pertinent. For example, a researcher working on a 1590 baptism register from Havana, does not require data contained in an 1820 marriage volume from the city of Limonar. The inclusion of extraneous materials in the visible lists of people and events would be a distraction and could lead to false detection of possible duplicate entries. Thus, the document hierarchy for the entirety of the *SSDA Cuba* project shown in Figure 5.13 indicates marriage records from Havana beginning in 1584 through burial records from Limonar with dates into the year 1892. At the time of this writing, the names of 1,517 individuals had been identified from this entire set of documents. By applying a filter, a researcher working on a single volume, *LIBRO DE BARAJAS: Bautismos 1590-1600* for

instance, would have the number of individuals listed reduced by over half and would still have a very high degree of confidence that they were distinct across the project.
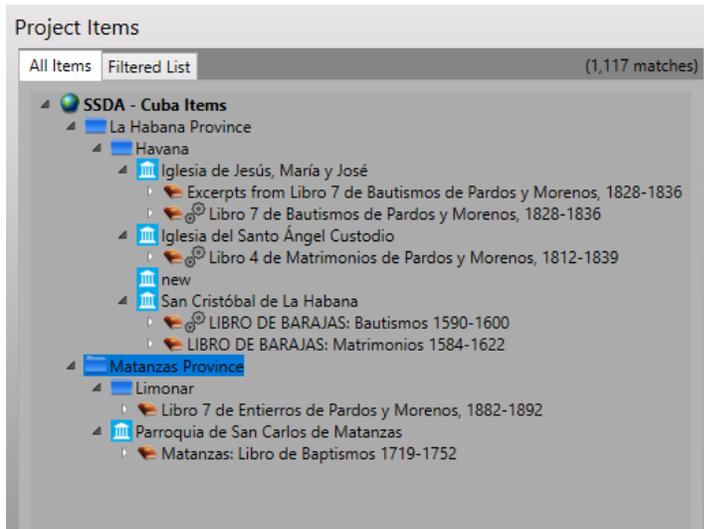
Figure 5.13: The SSDA Cuba document tree showing contents from two cities ranging date from 1584 through 1892.

The simplest way to apply a filter to the list of people is to select and then drag an item from the list of documents and drop it on the *People* window. This immediately adjusts the list of people to those contained in the selected item and all its subordinate items. So, for example, selecting the Havana node in Figure 5.12 and dropping it on the *People* window would cause the display to list all individuals identified in all the documents from the churches of *Iglesia de Jesús, María y José*, *Iglesia del Santo Ángel Custodio*, and *San Cristóbal de La Habana*. Likewise, dropping the item labeled *Folio 1 Recto* from *LIBRO DE BARAJAS: Bautismos 1590-1600* onto the *People* window would list the 19 individuals from that page and as shown in Figure 5.14.
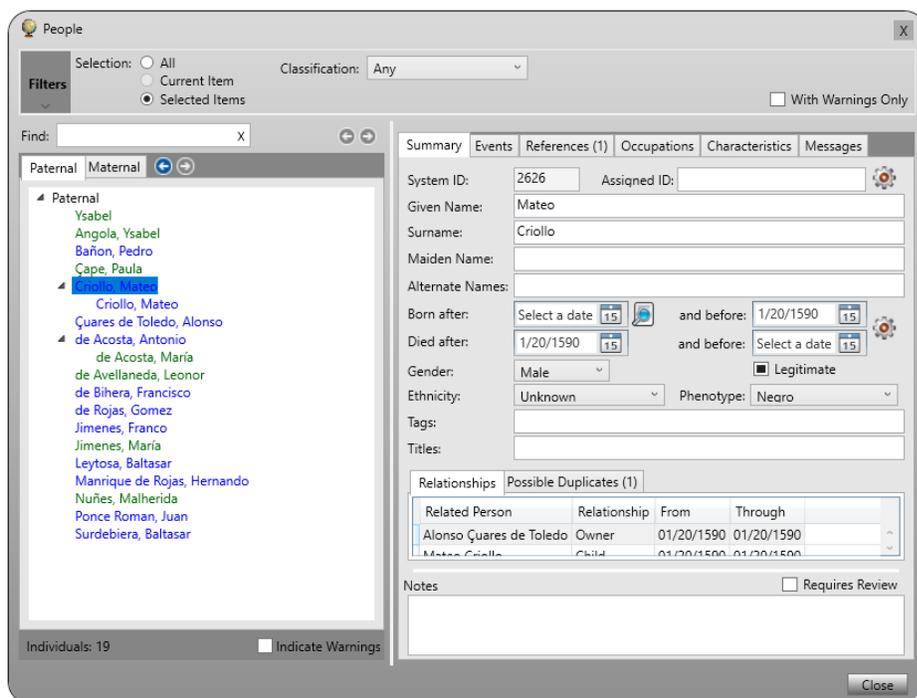
Figure 5.14: The People window with the list of individuals filtered to show only those identified on Folio 1 Recto of LIBRO DE BARAJAS: Bautismos 1590-1600.

The same filtering mechanism can be applied to the Events window which enables the user to focus on the events contained in a specified selection of items. Because of the tight coupling between the Event and the Map windows, the user can easily map the events as shown in Figure 5.15.
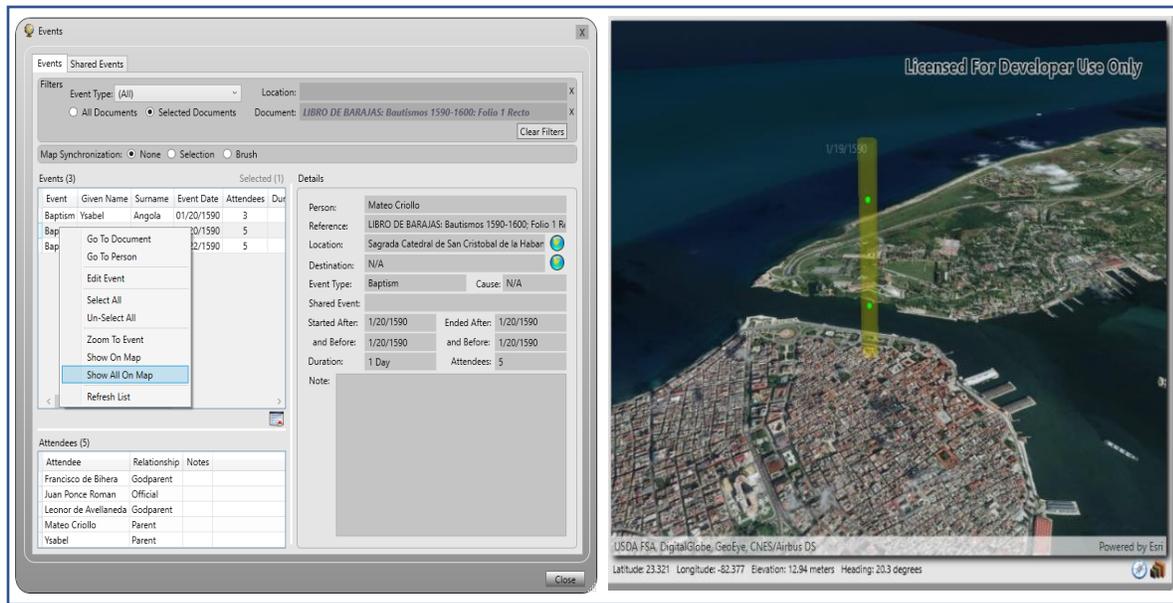
Figure 5.15: Example of mapping the filtered event list to show only those events identified on Folio 1 Recto of LIBRO DE BARAJAS: Bautismos 1590-1600.

.

Matrix File Export

The SSDA partnership in the *Enslaved: People of the Historic Slave Trade* project further required the ability to export the extracted content into a format specified by the Matrix team at MSU. This requirement drove the further development of the Spatial Historian's export feature which allows users to export information related to individual documents, individual people, or entire projects. Currently the system supports the export of data in the layout specified by the Matrix project as either an Excel spreadsheet or a comma separated value (CSV) file (Figure 5.16). In all cases, the system creates a folder containing individual files containing information with detailed information about source documents along with information about the individuals and places they reference. The exported data includes summary information about individuals along with life events, relationships, the places they have travelled, and a snapshot in time.
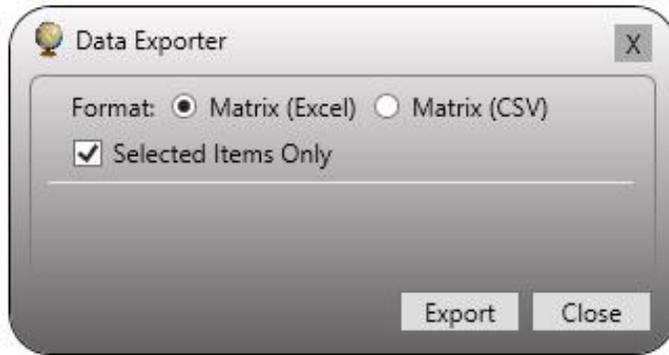
Figure 5.16: Form used to select the output format and the items to export.

The exported content is contained in a set of six files created under a folder generated specifically with each export request. The individual files are in MS Excel format (.xlsx files). Within the files unique identifiers are generated for all objects stored in the database to allow the MSU project team to reconstruct relationships and references. The identifiers can either be automatically generated by the system or manually assigned, and export files may contain combinations of the two. Manually assigned identifiers are used when available otherwise system assigned identifiers are written.

System generated identifiers are constructed using the format:

    <prefix>   - 'E': Event, 'I': Item, 'L': Location, 'P': Person, 'R': Reference, 'S': Snapshot
    NNN   - Project ID
    '-'
    NNNNNN - Internal identifier

Thus, identifier `P009-000701` would be assigned to the *Person* Jeronimo Martinez from the SSDA Cuba project who has the internal ID 701. Manually assigned identifiers are created based on individual project requirements such that York University's *Freedom Narratives* project (a contributor to the MSU Enslaved Hub) assigns IDs for people formatted with 'FN' followed by a 6-digit number, for example 'FN000005'.

Case Study Outcomes

The value of the Spatial Historian to historians has been clearly demonstrated for the SSDA project by the extensive use of the system by the project team to date. The baseline functionality that allows the names of the enslaved individuals to be recorded along with their basic demographic details has been utilized heavily to fulfill the goals of the SSDA project. In addition, functionality such as the event stack of baptisms within a church or group of churches for any period of time further opportunity for real-time analysis to help identify anomalies or discernible patterns. An examination using the Event viewing interface, for example, indicated that there was a change in one particular church representative who officiated over the baptismal ceremonies. Juan Ponce Roman had presided over the majority of the 103 baptisms recorded from January 20th, 1590 through January 30th, 1591. Following the ceremony recorded for January 30th his name no longer appears in the volume. Questions have been posed by project researchers as to whether he moved to a different church or returned to Spain, possibly died, or turned over his responsibilities to someone else. The fact that the last ceremony he presided over occurred at the end of a month and the subsequent ceremony occurred at the start of a month seems to indicate that the change was planned, or at least, organized in some way. From the perspective of the slave trade, an examination of other volumes recorded at the same church and those recorded in the region may prove useful in determining the causes of this type of change, their frequency, and the effect they may have had on the continuity of baptisms. In a broader context it may also prove beneficial as a source of information for scholars of the ecclesiastical history of the early years of the Catholic church in the Americas.

In similar vein other questions that arise from an examination of the interactions and relations between shared events and the life events of individuals that can be explored using the system functionality. For example, did the arrival of a slave ship at a local port precede an increased rate of baptisms among adult slaves in that newly arrived slaves were being baptized within some consistent timeframe of their arrival? Conversely, and possibly of greater import, could changes in the baptismal frequency of adult slaves indicate the ship on which individual slaves may have arrived on and thereby be used to develop an extended social network with other slaves on the same ship as well as the ship's captain and crew? These are tantalizing questions and are of considerable interest to the SSDA researchers (Pelegrino, 2019).

CHAPTER 6 - Challenges, Evaluation, Future Research, and Conclusion

Underpinning the Spatial Historian is the premise that adding geographical space into historical research and especially at the inception and initiation of a project is necessary and valuable. Insinuating space in historical research, however, is more than adding a mapping and spatial analysis capability to a platform as with a GIS, but in framing a system that is empathetic and adaptable to the unique needs of historians. Not least while the spatial turn has emphasized the importance of embedding a spatial perspective within the humanities, without the appropriate technological tools to do so makes such admonitions somewhat ephemeral and more wishful thinking than an actuality.

Both case studies confirmed the utility and value of a comprehensive and tightly integrated set of functional components available for the data capture, storage, management, and analysis of unstructured primary historical source materials. One of the existing impediments to the adoption of HGIS tools has been the steep technological learning curve needed to create viable integrated operational systems with the functional tools needed to support humanities and especially history scholars. Central to the uptake of such systems is not solely the platform functionality but providing an all-inclusive and intuitive interface that can be operationalized without extensive training and that can be customized for specific project needs to further develop knowledge and to generate additional questions arising from those insights. This capability is particularly beneficial when data from several projects is synthesized. In this way the extraction and digital organization of the data from historical documents offer many additional benefits that extend beyond the immediate objectives of an individual project. This benefit lies in the availability of the data for subsequent exploration and data mining by future generations of researchers. This becomes

particularly significant as additional data become accessible across multiple projects or, as in the case of SSDA, as additional volumes are transcribed and added to the collection. The "heads-down" transcription and day-to-day extraction activity leaves little time to consider aspects of the information content beyond the immediate objectives. However, the extracted content, with its rich spatial perspective and the social network linkages remain available for future analysis and do not themselves fall back into the chasm of *dark data*.

Project Goals and Questions Revisited

The goal of this dissertation was to develop, implement, and demonstrate an integrated customized geospatial system capable of assisting historians in the collaborative real-time capture and management of structured and unstructured data contained in archival source materials to facilitate subsequent data mining and geo-historical information exploration. Such a system, the Spatial Historian, has been developed to fulfill this role and has been tested in two ongoing historical case studies. To build such a system required a significant development effort comprised of 148,114 lines of code, 11,250 lines of XAML code, 101 XAML files, 157 database tables, 133 C# files, and 76 windows. A number of research questions were also identified to guide the study and its expected outcomes. Foremost among these was to conceptualize and develop an appropriate technological framework for addressing the challenges of working with unstructured historical materials. This conceptual technological framework was achieved through the development of a number of component parts that were integrated to form the Spatial Historian platform. Central to the platform was the creation of a flexible user interface and comprehensive data model that could facilitate the transcription, extraction, management, and exploration of information found in unstructured source materials. The early interface was further enhanced for the SSDA project

through the development of the Content Helper interface which has been recognized as a notable contribution to the SSDA project **Invalid source specified.**. The St. Vincent project benefitted from the system's ability to link toponyms to geographical locations, its ability to organize family genealogies, and its capacity to link people to places on the landscape. Contributing to the formation of the Spatial Historian conceptualization was the theoretical and operational work of the spatial humanities and deep mapping. These concepts contributed to an emphasis on a flexible user interface and the development of a comprehensive data model. The Spatial Historian offers an integrated solution with a powerful user interface as opposed to an off-the-shelf GIS software package or an amalgam of loosely couple software modules. Comparable functionality would require the use of a GIS for map production, a custom data model, a custom user interface for data entry, custom written logic for name disambiguation, a custom 3D spatial-temporal event visualization capability, a custom user interface to support the dynamic linking of people, places, events, and documents. As a deep mapping research tool, the Spatial Historian supports the collection of georeferenced documentary material and attachments that in combination promote a variety of deep mapping activities within spatial humanities projects.

The third research question sought to identify the needs and functionality required by historians to facilitate the transcription, normalization, and translation of historical texts. Part of these needs were identified through extensive discussions with historians and in particular with the principal investigators of the St Vincent and SSDA projects, partly through the initial prototype, and partly throughs a hands-on experience working on the case study projects. Through these means, and especially the case studies, an extensive set of practical requirements and functionality were identified and implemented. Because of the current state of OCR technology,

the transcription process within the Spatial Historian remains limited even though including a full-blown OCR and enhanced automated text processing would significantly improve the transcription and content extraction capabilities of the system. For both case studies, the ability to use a configurable synonym parser to standardize text and to match people and placenames has proven to be very effective. In the case of the SSDA project, support for transcription and content extraction is most notable in the creation of the Content Helper.

The fourth and fifth questions focused on the system functionality and database requirements needed to extract, store, and use spatial data and individuals taken from historical texts for assembly into spatialized family genealogies. This aspect is perhaps most clearly demonstrated in the St. Vincent case study where deep family trees were derived from the source materials and where ambiguity in the identity of individuals and spatial location was addressed through the system ability to link people with multiple individuals with similar names and through a toponym gazetteer and parser. The Spatial Historian evolved over the course of undertaking the two case studies which further exposed new system requirements and data linkages that resulted in enhancements to the database structure that support the comprehensive needs of the system.

A sixth question sought to identify the mechanisms required to facilitate the discovery, linking, and embedding of internet sources related to historical texts and to the people, places and events they reference. The answer to this is related to the deep mapping capabilities of the Spatial Historian. The Spatial Historian system enables users to attach external materials to the people, places, events, and documents it manages. These materials are incorporated via links to external items such as documents stored on a file system or images and pages on the internet. The external

materials can also be ingested directly into the Spatial Historians database from the actual sources. Linked materials have the advantage of being maintained and updated externally and thus their currency will always be up-to-date. One disadvantage to this is approach, however, is that the external items may be moved or deleted thereby rendering the referenced materials inaccessible. In contrast, ingested materials remain available to users of the Spatial Historian until they are explicitly removed though changes made to the original source materials will not be reflected in the database unless a call to the updated material is made. Materials attached to items within Spatial Historian can be viewed from within the system and can also function as supplemental materials for deep mapping projects.

A seventh question revolved around optimizing the user-interface not only for facilitating the extraction of historical content but also for interactive data exploration. The Spatial Historian was developed as a Microsoft Windows desktop application in order to take advantage of a development approach that is capable of tightly integrating and linking information across multiple windows and thereby allowing users to select events on one form that are then highlighted in the map interface. The desktop environment also supports drag-and-drop capability across multiple windows along with the ability to manage the system across windows. These advanced capabilities are difficult, if not, impossible to implement in a browser-based environment. However, because the system is a standalone desktop application, the software is required to be installed on each user's machine. A browser-based interface, while not as capable, could be loaded dynamically without the need for a formal installation process and there are additional stand-alone functions, such as the basic transcription process, that can be successfully implemented using a browser interface. Implementing a browser interface for transcriptions would then open up the opportunity

for crowdsourcing and greater throughput. In this manner also, the multi-tiered architecture of the Spatial Historian is capable of supporting additional user interface implementations beyond the desktop application.

An eighth concern focused on the functionality required to support the export of digital content for ingestion into archival and curated systems. The Spatial Historian addresses this question through its inherent collection of MODS compliant metadata data and through its export capability. The current export implementation supports the creation of files for ingestion into the MSU Enslaved data hub which is a set of export files constructed to a custom specification. These metadata can also be exported to other broadly recognized archival systems such a Fedora. However, because the metadata within the system is MODS compliant and because the export functionality is modular in design, implementing additional export file formats can be supported with only minor updates.

A final research question asks for a qualitative assessment based on feedback from users concerning the extent to which the system proved capable of assisting in the research objectives set out by these historians and in moving their historical research agenda forward. Such an evaluation includes ease of use, the comprehensiveness of the systems' capabilities, the system stability, desire to use the system, and overall satisfaction. In addition, the platform must be evaluated on how well the system can be deemed capable of extracting, integrating, and presenting the data in a manner that allowed the researchers to achieve their goals, to develop new historical insights, and, importantly, to open new lines of inquiry. In the case of the St. Vincent case study, this evaluation is somewhat more difficult to answer because while the technical functionality is in place, a more extensive application of the platform is required to evaluate its true effectiveness.

In the case of the SSDA case study it should be noted that the Spatial Historian is currently being used in production mode with more than ten students trained in its use. Plans are also currently in place at Vanderbilt University to integrate the system with their Fedora repository and to process additional document collections. The SSDA project also plan to enhance the system to make their data directly accessible through publishing REST services and a searchable web site that will support data downloads

System evaluation: The Baptism of Mateo Criollo Reprised

In several places in this dissertation, reference has been made to the January 20th, 1590 baptism of Mateo Criollo (Figure 6.1). The entry typifies the documentary information that can be derived from a seemingly unremarkable event and provides an example of how the content can be recorded, managed, and analyzed within the Spatial Historian. The record has additional instruction to offer about the value and use of the system as portrayed through the email exchange detailed below. In early August, Dr Jane Landers of Vanderbilt University was approached to submit an essay for publication in an edited volume for the Michigan State University *Enslaved: People of the Historic Slave Trade* project. She then invited Daniel Genkins and me to co-author. Again, the baptism of Mateo Criollo was used as a focal point for the discussion. This time, however, Genkins questioned the name that was transcribed as "Baltasar Surdebiera". The spelling, "Surdebiera", seemed to be inappropriate for a surname from the period, but the original document proved difficult to interpret (Figure 6.1). The authors decided to contact David Wheat, who performed the original transcription, to see if an alternative reading could be offered. The request and the resulting email correspondence are copied here**Invalid source specified.**. (All text and punctuation have been left verbatim)
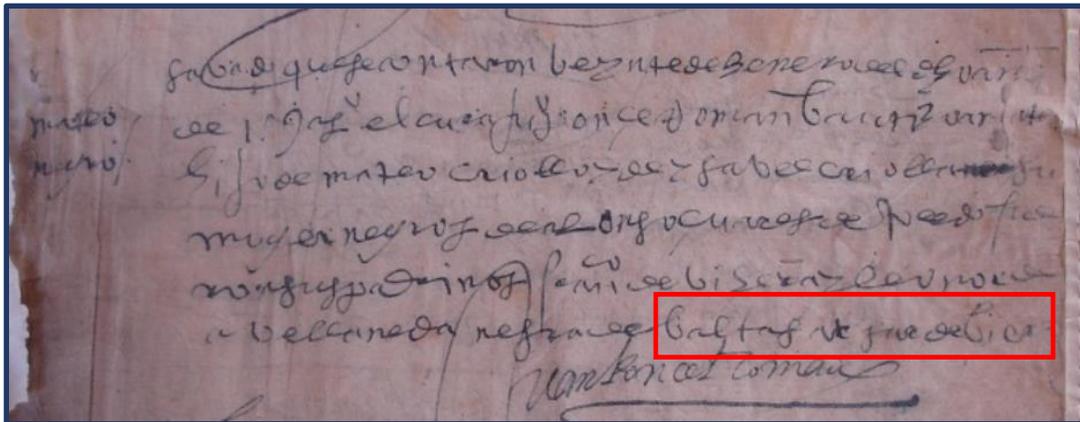
Figure 6.1: The baptism of Mateo Criollo with the name "Baltazar Surdebiera", aka "Baltazar de Biera", highlighted in the red rectangle.

Landers to Wheat

We are writing a chapter for MSU's Enslaved project...and wanted to use an example from the Libro de Barajas 1590 baptisms you worked in so long ago and transcribed...but you had a question by your own transcription and so I wanted to ask if you could take a look at the entry again...now that you've done so much more in the 16th c materials...and check it for us..

Wheat to Landers

I think it's "Baltasar sar de Biera." In other words I think the person writing messed up and wrote the last three letters a second time. The same guy shows up repeatedly as just "Baltasar de Biera" or "Baltasar de Bihera" on folios 66v, 69v, 106v, 111v, and 132v (in one case I have him as "de Bizera," oops!).

Schindling to Landers and Genkins

Interesting. So Baltasar may be related to Francisco de Bihera from the same entry?

Landers to Genkins and Schindling

Looks like

Mateo's madrina is Baltasar's slave..but Mateo must belong to Cuares as his parents do.. a...how Francisco fits is unclear, but if he were slave of Baltasar, it would say so like it does for the Madrina...maybe Baltasar' son? in household?

Genkins to Landers and Schindling

Actually, my reading now would be that both godparents were enslaved by Biera/Bihera. Wouldn't the convention it such a case be to write godfather A and godmother B slaves of X rather than godfather A slave of X and godmother B slave of X? Obviously the it should then read "negros" rather than "negra," but I think the easiest answer there is a little bit of sloppiness rather than a more complicated domestic situation!

Landers to Genkins and Schindling

Getting back to this...

I can read very little of the actual script but I think if Francisco were enslaved it would say "negro de" as it does for Leonor...

Wheat to Landers

Good luck with the essay, I'll look forward to reading it!
It would be awesome to try to reconstruct families among enslaved people in Havana households for that early period -- the marriage records might help, too, and Jorge's transcriptions are fantastic (and searchable!). Looks like Baltasar de Biera / Bihera and Beatriz Serrana ("Serana") show up several times; in addition to their own marriage they appear at least twice as godparents or slaveowners at slave weddings... maybe some useful info in there?
So far I think we're pretty lucky that there was this tradition of labeling Africans with "nations" -- I've been poking around in baptismal records for Santo Domingo during the 1590s and there's way less info on enslaved people, unfortunately...

Landers to Wheat

If you come down, or we up, or maybe via SKYPE you could also learn the Spatial Historian and it would help reconstruct networks of every kind much faster than we can individually.
And I think the "nations" and serially listing them over two centuries plus is the great contribution of SSDA
I have also been disappointed that in SD they didn't record them as meticulously as in Cuba and Fl.

The initial lesson to be learned from this correspondence thread is that transcription is not an easy task. When Genkins felt that the name "Surdebiera" had been transcribed incorrectly he looked at the image of the document yet, even though he himself has transcribed several volumes, he was unable to interpret the writing. Wheat, recently having worked with several similar texts, was asked to re-evaluate the text and was able to transcribe the surname more appropriately as "de Biera". This is an example of a common situation where although it is difficult to hire and train people to perform transcriptions, it is even more difficult to find people with significant levels of practical experience. The accuracy of the transcriptions across any corpus can vary widely based on who is performing the work. It is unlikely that OCR transcriptions would help for they require training datasets that are based on manually transcribed examples. When the training examples themselves are suspect, either due to faulty interpretation or errors in the original text, such as the extra "sar" in the case of the Mateo example, the process will inevitably produce inaccurate results.

However, the point to be made in this instance is that the Spatial Historian simultaneously enabled all involved to examine the original manuscript, the transcription of that entry, and everyone was able to compare the event with similar names and entries contained in the database. Once the name was re-interpreted, the discussion turned to the relationships between the individuals mentioned and their enslavement status. This in turn led to a further discussion of the different traditions of how the nationalities of slaves were documented at various locations in the Caribbean and the regional variation in the quality of the documentation. Finally, there is a mention in the email thread of using the Spatial Historian to help collaboratively reconstruct social networks. This is an arena where the software, applied as a tool, can help manage and analyze large volumes of data in ways that would not be possible if done manually. When applied appropriately, technology such as the Spatial Historian can aid in the identification and visualization of patterns and in the generation of new exploratory questions. In a twist on technology-meets-the-humanities, writing an essay about the development of a software system, rather than the software system itself, has generated a series of new discussions that may grow into new lines of research.

Future research

     Unlike software developed for many other dissertation projects, the Spatial Historian was intended from its inception to be a fully functioning production system. This vision and intent has shaped the development of many features that may otherwise have been set to the side, such as the *Content Helper*, text normalization, and event visualization. At the same time, this goal has required a significant amount of effort that might not have otherwise been undertaken. As of this writing, the system is use in at Vanderbilt University with its continued development funded for at least one additional year through a post-doc appointment. This will allow many envisioned enhancements to be implemented.

The Spatial Historian, as it now exists, was developed as user requirements were being defined and as programming techniques related to specific functionality were evolving. To avoid software licensing issues and to improve performance, the use of ESRI mapping services was replaced with custom coded graphics created using Microsoft SQL Server for spatial data storage. This evolutionary development approach required the adoption and learning of new technologies while simultaneously seeking to implement the required functionality. The result is a stable and functional system, yet one whose internal programming structure needs to be re-evaluated and made more modular.

Functionally the Spatial Historian has achieved all the goals that were set out to be achieved and the system has demonstrated its capability to assist in handling the sheer volume of documentary text and spatial toponyms that comprise unstructured archives and that make manual transcription and content extraction of the entire collection untenable. One significant operational benefit would be continued development of OCR capability and NLP capabilities. Research into the application of these technologies should be a priority over the near future in order to optimize the data transcription process. Data transfer standards, data curation, and the publication of the extracted content though clearinghouses and linked-open data sites such as the MSU Matrix Enslaved data hub will also continue to develop. Making data accessible by publishing REST services and a searchable web site that will contribute significantly to a greater sharing of these important datasets. Other identifiable enhancements revolve around the search capability of the system, or more appropriately the query and reporting capability, which is underdeveloped. Enhancements in this arena should be among the first to be implemented as new releases of the system are made. The impediments here are twofold: first the direct end-user querying of the database can be blocked for security reasons and a separate read-only reporting database for ad-

hoc querying and analysis is required. The second impediment is to move the reporting services to a cloud server rather than its current implementation running on a server housed in the WVGIS Technical Center. Because the server is behind the university firewall it cannot be accessed by users who do not have access to the WVU network and additional hosting fees would be required to support this move.

The research set out to produce a technology capable of supporting deep mapping and the work of historians mired in the dark data of unstructured archival materials. It is suggested here that the functionality provided by this system is a significant step forward in that it provides a capability to assist in pursuing a historian's research objectives. As Daniel Genkins of Vanderbilt University stated regarding the use of the Spatial Historian in support of the SSDA, the extraction of structured data from the unstructured texts is what is most important to their project**Invalid source specified.**. While the ability to extract and manage documentary archival content is indeed a large part of the Spatial Historian, there is much more potential to the system that has yet to be seen.

REFERENCES

Abbott, D. (2008). What is digital curation? *Digital Curation Centre*.

Alex, B., Byrne, K., Grover, C., & Tobin, R. (2015). Adapting the Edinburgh Geoparser for Historical Georeferencing. *International Journal of Humanities and Arts Computing, 9*(1), 15-35.

Alves, D., & Queiroz, A. (2015). Exploring Literary Landscapes: From texts to spatiotemporal analysis through collaborative work and GIS. *International Journal of Humanities and Arts Computing, 9*(1), 57-73.

Assessorato Territoria e Ambiente Valle d'Aosta. (2014, November 30). *Prodotti.* Retrieved from Sistema delle Conoscenze Territoriali: http://geoportale.partout.it/distribuzione_prodotti/cartografici/default_i.aspx

Ayers, E. L. (2010). Turning towards place, space, and time. In *The Spatial Humanities: GIS and the Future of Humaniities Scholarship* (pp. 1-13).

Berman, M. (2009). Modeling and visualizing historical gis data. *Proceedings of the Spatio-Temporal Workshop.*

Berman, M., Mostern, R., & Southall, H. (Eds.). (2016). *Placing Names : Enriching and Integrating Gazetteers.* Bloomington: Indiana University Press.

Bishr, Y. (2008). Geospatial Semantic Web: Applications. In *Encyclopedia of GIS* (pp. 391-398). Springer US.

Bleakly, D. R. (2002). Long-Term Spatial Data Preservation and Archiving: What are the issues? *Sand Report, 107*.

Bleakly, D. R. (2002). *Long-Term Spatial Data Preservation and Archiving: What are the Issues?* Sandia National Laboratories.

Bodenhamer, D. J. (2008). History and GIS: Implications for the Discipline. In A. K. Knowles (Ed.), *Placing History: How Maps, Spatial Data, and GIS are Changing Historical Scholarship* (pp. 219-233).

Bodenhamer, D. J. (2010). The Potential of Spatial Humanities. In D. J. Bodenhamer, J. Corrigan, & T. M. Harris (Eds.), *The Spatial Humanities: GIS and the Future of Humanities Scholarship* (pp. 14-30). Bloomington: Indiana University Press.

Bodenhamer, D. J. (2015). Narrating Space and Place. In D. J. Bodenhamer, J. Corrigan, & T. M. Harrris (Eds.), *Deep Maps and Spatial Harratives* (pp. 7-27). Bloomington: Indiana University Press.

Bodenhamer, D. J., Corrigan, J., & Harris, T. M. (Eds.). (2010). *The Spatial Humanities: GIS and the Future of Humanities Scholarship.* Bloomington: Indiana University Press.

Bodenhamer, D. J., Corrigan, J., & Harris, T. M. (Eds.). (2015). *Deep Maps and Spatial Narratives.* Bloomington: Indiana University Press.

Bose, R., & Reitsma, F. (2005). Advancing Geospatial Data Curation. *International Journal of Digital Curation.*

Braudel, F. (1995). *The Mediterranean an the Mediterranean world in the age of Philip II* (Vol. 1). University of California Press.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). *Digital_Humanities.* Cambridge: The MIT Press.

Caplan, P. (2009). *Understanding PREMIS.* Library of Congress.

Castree, N. (2005). Is Geography a Science. In N. Castree, A. Rogers, & D. Sherman (Eds.), *Questioning Geography: Fundamental Debates* (pp. 57-79). Malden, MA: Blackwell Publishing.

CCSDS. (2012). *Reference Model for an Open Archival Information System (OAIS).* The Consultative Committee for Space Data Systems. Retrieved from https://public.ccsds.org/Pubs/650x0m2.pdf

Digitisation and Digital Preservation group. (2018, August 1). *Transkribus.* (University of Innsbruck) Retrieved from Transkribus: https://transkribus.eu/Transkribus/

Dragicevic, S., Marceau, D., & Marois, C. (2001). Space, time, and dynamics modeling in historical GIS databases: a fuzzy logic approach. *Environment and Planning B: Planning and Design, 28*(4), 545-562.

Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly, 5*(1), 1-21.

ESRI. (2019, August 1). *ArcGIS Pro.* Retrieved from Use lidar in ArcGIS Pro: https://pro.arcgis.com/en/pro-app/help/data/las-dataset/use-lidar-in-arcgis-pro.htm

Ethington, P. J. (2007). Placing the Past: 'Groundwork' for a Spatial Theory of History. *Rethinking History, 11*(4), 465-493.

Fifield, W. (1964). Pablo Picasso: A composite interview. *The Paris Review 32. 1964, 37.*

Foucault, M. (1984). *The Foucault Reader.* (P. Rabinow, Ed.) Vintage Books.

Frenkel, E. (2014, February 14). Ad Infinitum: 'Our Mathematical Universe' by Max Tegmark. *Sunday Book Review.*

Garrison, T. G., Houston, S., & Firpi, O. l. (2019). Recentering the rural: Lidar and articulated landscapes among the Maya. *Journal of Anthropological Archaeology, 53*, 133-146.

Goodchild, M. F. (1992). Integrating GIS and spatial data analysis problems and possibilities. *International Journal of Geographical Information Systems, 6*(5), 407-423.

Goodchild, M. F. (2004). GIScience, Geography, Form, and Process. *Annals of the Association of American Geographers, 94*(4), 709-714.

Goodchild, M. F. (2006). Geographical Information Science Fifteen Years Later. *Classics from IJGIS: twenty years of the Internatinal Journal of Geographical Information Science and Systems*, 199-203.

Goodchild, M. F. (2008). Combining Space and Time: New Potential for Temporal GIS. In A. K. Knowles, & A. Hiller (Eds.), *Placing History: How Maps, Spatial Data, and GIS are Changing Historical Scholarship* (pp. 179-197). Redlands: ESRI Press.

Google. (2019, August 1). *Cloud Translation Documentation*. Retrieved from AI & Machine Learning Products: https://cloud.google.com/translate/docs/

Gregory, I. (2002). Time-variant GIS databases of changing historical administrative boundaries: A European comparison. *Transactions in GIS, 6*(2), 16-178.

Gregory, I. N., & Ell, P. S. (2007). *Historical GIS: Technologies, Methodologies, and Scholarship.* New York: Cambridge University Press.

Gregory, I. N., & Healey, R. G. (2007). Historical GIS: structuring, mapping and analysing geographies of the past. *Progress in Human Geography, 31*(2), 638–653.

Gregory, I. N., Donaldson, C., Murrieta-Flores, P., & Rayson, P. (2015). Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research. *International Journal of Humanities and Arts Computing, 9*(1), 1-14.

Gregory, I., Cooper, D., Hardie, A., & Rayson, P. (2015). Spatializing and Analysing Digital Texts: Corpora, GIS, and Places. In D. J. Bodenhamer, J. Corrigan, & T. M. Harris (Eds.), *Deep Maps and Spatial Narratives* (pp. 150-178). Bloomington: Indiana University Press.

Guelke, L. (1997). The Relations between Geography and History Reconsidered. *History and Theory, 36*(2), 216-234.

Guldi, J. (2017). *Spatial Humanities: What is the Spatial Turn?* Retrieved from Scholars Lab, University of VIrginia Library: http://spatial.scholarslab.org/spatial-turn/what-is-the-spatial-turn

Hägerstraand, T. (1970). What about people in regional science? *Papers in regional science, 24*(1), 7-24.

Hagerstrand, T. (1970). Regional forcasting and social engineering. *Regional forcasting*, 1-7.

Harris, T. (2015). Deep Geography - Deep mapping: Spatial story telling and a sense of place. In D. J. Bodenhamer, J. Corrigan, & T. M. Harris (Eds.), *Deep Maps and Spatial Narratives* (pp. 28-53). Bloomington: Indiana University Press.

Harris, T. M. (2002). GIS in Archaeology. In K. A. Knowles, *Past Time, Past Place: GIS for History* (pp. 131-143). Redlands: ESRI Press.

Harris, T. M. (2017). Deep mapping and sensual immersive geographies. In *The International Encyclopedia of Geography*.

Harris, T. M. (2017). Deep mapping and sensual immersive geographies. In *The International Encyclopedia of Geography*.

Harris, T. M., Bergeron, S., & Rouse, L. J. (2011). Place, spatial storytelling, and immersive visualization in the humanities. In M. Dear, *Geohumanities: Art, History, Text at the Edge of Place* (pp. 226-240).

Harris, T. M., Rouse, L. J., & Bergeron, S. (2010). The Geospatial Semantic Web, Pareto GIS, and the Humanities. In *The Spatial Humanities: GIS and the Future of Humanities Scholarship* (pp. 124-142).

Harvey, D. (1990). Between Space and Time: Reflections on the Geographical Imagination. *Annals of the Association of American Geographers, 80*(3), 418-434.

Harvey, D. (2001). *Spaces of Capital: Towards a Critical Geography*. Edinburgh: Edinburgh University Press.

Harvey, D. W. (1967). Gehavioural Postulates and the Construction of Theory in Huuman Geography. *Studies in Geographical Methods* (pp. 27-45). Warsaw: PWN.

Harvey, D. W. (1974). What Kind of Geography for What Kind of Public Policy? *Transactions of the Institute of British Geographers, 63*, 18-24.

Harvey, D. W. (1984). On the History and Present Condition of Geography: An Historical Materialist Manifesto. *The Progessional Geographer, 36*(1), 1-11.

Heidorn, P. B. (2008). Shedding Light on the dark data in the long tail of science. *Library Trends, 57*(2), 280-299.

HTC Inc. (2019, August 1). *Products*. Retrieved from vive: https://www.vive.com/us/

Illusion of the Year. (2017). *Best Optical Illusion of the Year Contest*. Retrieved from Illusion of the Year: http://illusionoftheyear.com/2017/10/shape-from-motion-only/

International Association of Oil and Gas Producers. (2019, August 1). *About the EPSG Dataset*. Retrieved from International Association of Oil and Gas Producers: http://www.epsg.org/

Jessop, M. (2004). The Visualization of Spatial Data in the Humanities. *Literary and Linguistic Computing, 19*(3), 335-350.

Jordan, T. (1990). The intellectual core of Geography was endangered by "the rush to GIS and similar easily justified but non-intellectual expertise". *AAG Newsletter, 23*(5).

Kennan, M., & Markauskaite, L. (2015). Research Data Management Practices: A Snapshot in Time. *The International Journal of Digital Curation, 10*(2), 69-95.

Knowles, A. K. (2000). Special Issue: The spatial turn in social science history. *Historical GIS*, 451-470.

Knowles, A. K. (2005). Emerging trends in historical GIS. *Historical Geography, 33*, 7-13.

Knowles, A. K. (2008). GIS and History. In A. K. Knowles (Ed.), *Placing History: How Maps, Spatial Data, and GIS are Changing Historical Scholarship* (pp. 1-26). ESRI Press.

Knowles, A. K. (2008). What Could Lee See at Gettysburg? In A. K. Knowles (Ed.), *Placing History: how maps, spatial data, and GIS are changing historical scholarship* (pp. 235-265). Redlands: ESRI Press.

Knowles, A. K. (2014). The contested nature of historial GIS. *International Journal of Geographical Information Science, 28*, 206-211.

Knowles, A. K. (2016). Historical Geographic Information Systems and Social Science History. *Social Science History, 40*(4), 741-750.

Knowles, A. K., & Hillier, A. (Eds.). (2008). *Placing History: how maps, spatial data, and GIS are changing historical scholarship.* Redlands: ESRI.

Knowles, A. K., Westerveld, L., & Strom, L. (2015). Inductive visualization: A humanistic alternative to GIS. *1*(2), 233-265.

Kong, N. (2016). The Geospatial Data Curation, Management, and Discovery in Academic Libraries. *IFLA World Library and Iincormation Congress* (pp. 1-12). Columbus: IFLA.

Landers, J. (2019, April 1). Personal Interview. (J. Schindling, Interviewer)

Lefebvre, H. (1991). *The Production of Space* (Vol. 142). Oxford: Blackwell.

Lipe, W. D. (2000). Conserving the in situ archaeological record. *Conservation Perspectives, The GCI Newsletter, 15*.

Lord, P., Macdonald, A., Lyon, L., & Giaretta, D. (2004). From Data Deluge to Data Curation. *Proceedings of the UK e-science All Hands meeting*, (pp. 371-375).

Manguinhas, H., Martins, B., Borbinha, J., & Siabato, W. (2009). The DIGMAP Geo-Temporal Web Gazetteer Service. *e-Perimetron, 4*(1), 9-24.

Marshall, P. (2007). Do tangible interfaces enhance learning? *1st International conference on tangible and embeded interaction* (pp. 163-170). Baton Rouge: The Open University.

Martins, B., Anastácio, I., & Calado, P. (2010). A machine learning approach for resolving place references in text. *Geospatial Thinking*, 221-236.

Martins, B., Manguinhas, H., & Borbinha, J. (2008). Extracting and exploring the geo-temporal semantics of textual resources. *2008 IEEE International Conference on Semantic Computing* (pp. 1-9). IEEE.

McGarva, G., Morris, S., & Janee, G. (2009). *Technology Watch Report: Preserving Geospatial Data*. Dgital Preservation Coalition. Retrieved from http://www.dpconline.org/advice/technology-watch-reports

McGlashan, C. F. (1947). *History of the Donner Party: A Tragedy of the Sierra*. Stanford University Press.

Metadata Ad Hoc Working Group. (1998). *Content Standard for Digital Geospatial Metadata*. Washington, DC: Federal Geographic Data Committee.

Micorsoft. (2018, January 1). *Microsoft Azure Documentation*. Retrieved from Microsoft Azure: https://docs.microsoft.com/en-us/azure/#pivot=products&panel=cognitive

Microsoft. (2018, January 1). *Get started with Speech Service client libraries*. Retrieved from Microsoft Azure: https://docs.microsoft.com/en-us/azure/cognitive-services/speech/getstarted/getstartedclientlibraries

Microsoft. (2018, January 1). *Image Search API v7 reference*. Retrieved from Microsoft Azure: https://docs.microsoft.com/en-us/rest/api/cognitiveservices/bing-images-api-v7-reference

Moore, R. W., Rajasekar, A., & Marciano, R. (2007). Implementing Trusted Digital Repositories. *DigCCurr2007 International Symposium on Digital Curation, UNC*.

Moyes, H., & Montgomery, S. (2019). Locating Cave Entrances Using Lidar-Derived Local Relief Modeling. *geosciences, 9*(2).

Mueller III, R. S. (2019). *Report On The Investigation Into Russian Inerference In The 2016 Presidential Election*. Washington DC: U.S. Department of Justice.

Nickens, P. R. (2000, August 20). *The destruction off archaeological sites and data*. Retrieved March 5, 2013, from National Park Service: http://www.cr.mps.gov/seac/protecting/html/3a-nickens.htm

NISO. (2004). *Unserstanding Metadata*. Bethesda: NISO Press.

OCLC/RLG Working Group on Preservation Metadata. (2002). *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects*. Dublin, Oh: Online Computer Library Center.

Oculus. (2019, August 1). *Oculus*. Retrieved from Oculus: https://www.oculus.com/

Open Geospatial Consortium. (1999). Consortium: OpenGIS Simple Features Specification for SQL, Revision 1.1. *OpenGIS Project Document*.

Open Geospatial Consortium. (2010). OpenGIS Implementation Specification for Geographic information-Simple feature access-Part 2: SQL option. *OpenGIS Implementation Standard*.

Open Geospatial Consortium. (2011). OpenGIS® Implementation Standard for Geographic information-Simple feature access-Part 1: Common architecture. *OpenGIS Implementation Standard*.

Open Geospatial Consortium. (2017). *OGC*. Retrieved from OGC: http://www.opengeospatial.org/

Owens, J. B. (2007). Towards a geographically-integrated, connected world history: Employing geographic information systems (GIS). *History Compass, 5*(6), 2014-2040.

Owens, J. B. (2007). Towards a geographically-integrated, connected world history: Employing geographic information systems (GIS). *History Compass, 5*(6), 2014-2040.

Pavlovskaya, M. (2006). Theorizing with GIS: A tool for critical geographies. *Environment and Planning, 38*(11), 2003-2020.

Pearson, M., & Shanks, M. (2001). *Theatre / Archaeology*. London: Routledge.

Pleiades Individual Contributors. (2018, August 1). *Pleiades*. (S. C. Ancient World Mapping Center, Producer) Retrieved from Pleiades: https://pleiades.stoa.org/

Regione Autonoma ceila Calle D'Aosta. (2007). Provincia di Torino Comune di S. Vincent: 1895-1897.

RLG - OCLC. (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. Research Libraries Group - Online Computer Library Center. Retrieved from http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf

RLG Task Force on Archiving of Digital Information. (1996). *Preserving Digital Information*. Research Libraries Group.

Roman Roads Research Association. (2018, August 1). *The Roads of Roman Britain*. Retrieved from Roman Roads Research Association: http://roadsofromanbritain.org/index.html

Rothenberg, J. (1999, February 22). *Ensuring the Longevity of Digital Information*. Retrieved from CLIR: https://www.clir.org/pubs/archives/ensuring.pdf

Santos, J., Anastacio, I., & Martins, B. (2015). Using machine learning methods for disambiguating place references in textual documents. *Geojournal, 80*(3), 375-392.

Schindling, J. (2016). Challenging Historical GIS through Unstructured Historical Primary Source Materials: Uncovering Sixteenth-Century St. Vincent, Aosta, Italy. *Masters Thesis*.

Schindling, J., & Gibbes, C. (2014). LiDAR as a tool for archaeological research: a case study. *Archaeological and Anthropological Sciences, 6*(4), 411-423. doi:10.1007/s12520-014-0178-3

Schlichting, K. (2008). Historical GIS: New ways of doing history. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 41*(4), 191-196.

Smail, D. L. (1999). *Imaginary Cartographies: Possession and Identity in Late Medieval Marseille.* Ithaca: Cornell University Press.

Smith, D. A., & Crane, G. (2001). Disambiguating geographic names in a historical digital library. *In Research and Advanced Technology for Digital Libraries*, 127-136.

Soja, E. W. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory.* London: Verso.

Southall, H. (2003). Defining and identifying the roles of geographic references within text: Examples from the Great Britain Historical GIS project. *In Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references-Volume 1* (pp. 69-78). Association for Computational Linguistics.

Southall, H. (2012). Rebuilding the Great Britain Historical GIS, Part 2: a geo-spatial ontology of administrative units. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 45*(3), 119-134.

Southhall, H. (2011). Rebuilding the Great Britain Historical GIS, Part 1: Building an indefinitely scalable statistical database. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 44*(3), 149-159.

Springett, S. (2015). Going Deeper or Flattening: Connectin Deep Mapping, Flat Ontologies and the Democratization of Knowledge. *Humanities, 4*(4), 623-636.

Stegner, W. E. (1992). *The Gathering of Zion: The Story of the Mormon Trail.* Lincoln: University of Nebraska Press.

Stolze, K. (2003). SQL/MM Spatial - The Standard to Manage Spatial Data in a Relational. *BTW*, (pp. 246-264).

Strassler, R. B. (Ed.). (2009). *The Landmark Herodotus: The Histories.* (A. L. Purvis, Trans.) New York: Random House LLC.

Tegmark, M. (2014). *Our Mathematical Universe: My quest for the ultimate nature of reality.* Penguin UK.

TEI. (2018, January 1). *TEI: Text Encoding Initiative.* Retrieved from TEI: Text Encoding Initiative: www.tei-c.org/index.xml

The Polis Center. (2017). *Virtual Center for Spatial Humanities.* Retrieved from Spatial Humanities: http://polis.iupui.edu/index.php/spatial-humanities/project-1/

Tobler, W. R. (1970). A computer movie simulating urban growth. *Economic Geography, 46*(2), 234-240.

Tuan, Y.-F. (1975). Images and Mental Maps. *Annals of the Association of American Geographers, 65*(2), 205-213.

Tuan, Y.-F. (1976). Humanistic Geography. *Annals of the Association of American Geographers, 66*(2), 266-276.

Tuan, Y.-F. (1977). *Space and Place: The Perspective of Experience.* Minneapolis, MN: University of Minnesota Press.

Tuan, Y.-F. (1979). Space and place: Humanistic perspective. *Philosophy in geography*, 387-427.

U.S. Geological Survey. (2014, November 30). *Welcome to the USGS Center for LIDAR Information Coordination and Knowledge*. Retrieved from Center for LIDAR Information Coordination and Knowledge: http://lidar.cr.usgs.gov

Unity Technologies. (2018, January 1). *Unity User Manual (2017.3)*. Retrieved from Unity Documenation: https://docs.unity3d.com/Manual/index.html

Valve Corporation. (2018, January 1). *Steam*. Retrieved from Steam: http://store.steampowered.com/

Valve Corporation. (2019, August 1). *Command Line Options*. Retrieved from Valve Developer Community: https://developer.valvesoftware.com/wiki/Main_Page

Vanderbilt College of Arts and Sciences. (2019, April 1). Retrieved from Vanderbilt Center for Digital Humanities: https://www.vanderbilt.edu/digitalhumanities/

Vanderbilt University. (2018, April 1). *Transcriptions and Volume Descriptions*. Retrieved from Slave Societies Digital Archive: https://www.vanderbilt.edu/esss/cuba/transcriptions/transcriptions.php

Vanderbilt University College of Arts and Science, Department of History. (2018, April 1). *Slave Society Digital Archive*. Retrieved from Slave Society Digital Archive: https://www.vanderbilt.edu/esss/index.php

Vester, M. (2014, November). Personal Interview. (J. Schindling, Interviewer)

Vester, M. (2019, July 9). Personal Correspondence.

Viles, H. (2005). A Divided Discipline? In N. Castree, A. Rogers, & D. Sherman (Eds.), *Questioning Geography: Fundamental Debate* (pp. 26-38). Blackwell Publishing.

Visbox. (2019, August 1). *CAVE Automatic Virtual Environment*. Retrieved from Visbox: http://www.visbox.com/products/cave/

Walters, T., & Skinner, K. (2011). *New Roles for New Times: Digital Curation for Preservation.* Washington, DC: Association of Research Libraries.

Wernke, S. A., Adams, J. A., & Hooten, E. R. (2014). Capturing Complexity: Toward an Integrated Low-Altitude Photogrammetry and Mobile Geographic Information System Archaeological Registry System. *Advances in Archaeological Practice, 2*(3), 147-163.

White, R. (2010). *What is spatial history. Spatial History Lab: Working paper.* Retrieved from Stanford University Spatial History Project: http://www. stanford. edu/group/spatialhistory/cgi-bin/site/pub. php

Wood, D. (2016). Mapping Deeply. *Humanities, 4*, 304-318.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* PublicAffairs.

APPENDIX A

Demonstration Videos

The following videos are designed to provide an overview of the features and functionality available in the Spatial Historian user interface. The videos can be viewed in any order; however, it is recommended that they be viewed in the order they appear as references may be made to previously discussed functionality.

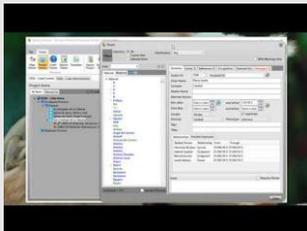These videos are also available on the Spatial Historian YouTube channel at:
https://www.youtube.com/channel/UCWnMZmTMyIEeO7V3qFuEZsQ
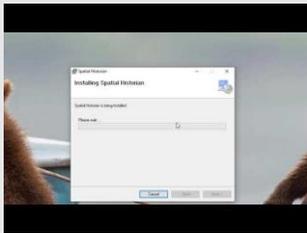
- CLIP 1 – 2:56 -
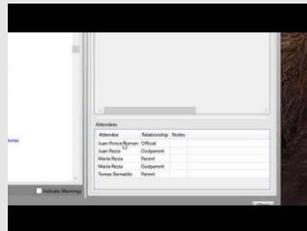INSTALLING SPATIAL HISTORIAN

- CLIP 2 – 3:09 -
LOGGING IN AND OPENING A PROJECT

- CLIP 3 - 0:54 -
MANAGING DOCUMENTS

- CLIP 4 – 12:44 -
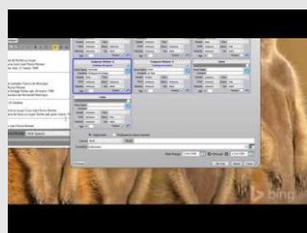THE PEOPLE VIEWER PART 1

- CLIP 5 - 0:54 -
THE PEOPLE VIEWER PART 2

- CLIP 6 - 0:54 -
THE MAP VIEWER

- CLIP 7 - 0:54 -
THE EVENT VIEWER

- CLIP 9 - 13:48 -
THE CONTENT HELPER

APPENDIX B

User Comments and Feedback

Dr. Matthew Vester

The goal and challenges of the St. Vincent are articulated in the following statement provided by Dr. Vester. This was used as both the driving vision for project specific enhancements and as the defining criteria for evaluating the success of the system.

(August 2019)

The goal of this project is to spatialize kinship relations within a rural mountain community during the early modern period, in order to compare family dynamics within and between rural families (in small towns and villages) with those in other social settings (in cities, amongst nobles and elites, and even within and between ruling dynasties). An historical GIS is a crucial element of this research, since it permits the research to document the spatial component of families' interests (locations of residences, properties, and landed interests) and measure the impact of this spatial dimension on family members' behaviors.

Historians of the early modern period always confront limitations in their primary sources. The St Vincent case is no different. The records used to locate family interests (feudal records, notarial records, and some parish records) are often temporally discontinuous. Even when they are continuous (such as in the notarial records, which provide most of the evidence for this project) they do not provide a complete picture of families' property interests. For example, while there are land sales, rental agreements, underwriting arrangements, marriage contracts with dowry information, property divisions, and wills, not every interaction was documented by a notary, and some families may have used notaries whose records did not survive. In addition to this, there are two other chief obstacles to achieving this project's goals, both of which deal with uncertainty in interpreting the data. One is the difficulty in determining whether individuals identified in various records who have the same name (and perhaps the same father's name) are in fact the same person, or whether they are different people. A second obstacle concerns the spatialization of family property and linking the spatial references in the sources to places that can be mapped in a GIS. The project has attempted to do this by associating the topoynms mentioned in the sources (every property mentioned in the notarial records was located within a toponym written in Valdotain – a form of Franco-Provençal) with the toponyms listed (in Italian) in a late nineteenth-century cadastral map. The latter permits

the researcher to geo-reference the nineteenth-century toponyms. The challenge that remains is matching the sixteenth-century toponyms to the nineteenth-century ones.

The uncertainty implicit in these two obstacles can be mitigated in some circumstances, but (in all likelihood) never completely eliminated. Regarding the names of individuals (such as whether two documents that identify a Laurent, son of Antoine who lived during the same period were in fact the same person), some certainty might be established by continuing to read through all available notarial records to see whether others records refer to given individuals in a way that permits confirmation that the names in the record do or do not refer to the same person. This would require funding (for graduate researchers to complete this work expeditiously), or more time (for the project leader himself to work slowly through the material as he carries out his other duties). The uncertainty about the relationship between the sixteenth-century Franco-Provençal toponyms and the nineteenth-century Italian ones could be mitigated through working with linguistic authorities in the Valle d'Aosta (at the Bureau régional pour l'ethnologie et la linguistique in Aosta and the Centre d'études francoprovençales in St Nicolas) and through oral interviews with inhabitants of St Vincent and its various villages.

Thus, through some additional work, some of the uncertainty associated with this project can be reduced, even though some questions (about individual identities and spatial references) will probably remain open due to the structural limitations of the evidence.

Alexandre Pelegrino

Alexandre Pelegrino is a doctoral candidate at Vanderbilt University. He uses the Spatial Historian as part of the SSDA project. He is interested in Latin American and Native American history with particular emphasis on the enslavement of indigenous people in colonial Brazil. His advisor is Professor Jane Landers.

(August 2019)

"The goal of my project is to understand how the increasing levels of market integration changed Maranhão`s [Brazil] society. Up until the mid-eighteenth century, the region was a periphery in the Portuguese empire, where settlers developed ranching activities with overwhelmingly Native workforce, enslaved and coerced. In the 1750s, however, the imperial reforms not only abolished Indian slavery, but the Portuguese crown also created a monopolistic trade company to ship an unprecedented number of African slaves to develop a cash crop economy based on cotton and rice production. There are two questions that I want to answer. First, I want to overcome the interpretations that understand Indian slavery as a practice confined to the early moments of colonization. By following the Indians` lives, I can draw a picture of their inclusion and permanence in households. Second, and related to that, I want to question if masters included Africans and Natives differently in their households by following patterns in parish records.

"The bedrock of my approach is a database combining six different serial primary sources (baptismal records, marriage records, burial records, notary records, wills, and land grants). Before starting using Spatial Historian, I was working on Excel spreadsheets, which generated several problems. Because each book had its own spreadsheet, I had to manually track down people over the different spreadsheet, a method that would not be feasible with the growth of the database. Using Spatial Historian, I can more easily aggregate different information for each person and avoid the danger of duplicates. I believe that the software will let me track families over more than one generation, allowing the visualization of how people changed classifications over time. One example of "social ascension." On June 29, 1748, the soldier Manoel do Nascimento (id 4417) married Claudina Rodrigues. While he was the son of Pio Nunes and Antonia Amada, his wife was the daughter of Antonio Rodrigues and the índia Joana. It is worth noting that neither Joana`s daughter nor her grandson, Joaquim, baptized on June 3, 1758, received the classification "índio" in their sacramental registers.

"My major problem with Spatial Historian is the data entry process. When I was working with spreadsheets, the process was much quicker. Another problem is the lack of more robust search capabilities, such as by occupation and by status. Currently, the software only allows filtering by classification."

Dr. Daniel Genkins

Daniel Genkins is a post-doctoral researcher at Vanderbilt University interested in digital humanities with historical applications focused on colonial Latin America and the early modern Atlantic World. His dissertation explored Anglo-Spanish encounters in the seventeenth-century Caribbean. His paper "'To Seek New Worlds, for Gold, for Praise, for Glory': El Dorado and Empire in Sixteenth-Century Guiana," which serves as a point of departure for the dissertation, received the Moseley Award for best graduate paper at the 2013 meeting of the Southeastern Council of Latin American Studies and was subsequently published in The Latin Americanist.

(August 2019)
"The SSDA use case for Spatial Historian was/is fairly specific: extracting structured information from manual transcriptions of digital images to contribute to the Enslaved project. Although this did not initially align exactly with the functionalities of SH, it was useful for our purposes from the start and became more so over time as Jim developed the tool with our needs in mind.

"The single most useful graphical element has probably been the "Content Helper" interfaces, which facilitate the extraction of information from certain kinds of records (e.g. baptisms, marriages, etc.), streamlining the data entry process. Other particularly useful pieces include a text "normalization" tool that converts abbreviations and archaic spellings to modern Spanish, a matching function that allows individuals to be disambiguated, and a visualization interface that allows events to be mapped over time and space. In addition, the experience of working with Jim to think through what SSDA needed the tool to do helped us to develop the controlled vocabularies that we use to populate some fields in our data.

"There are *many* aspects of SH that SSDA has not exploited to the fullest, most notably the aforementioned mapping capabilities. This is not because they're not useful or robust (in fact, they're likely among the most advanced tools that SH has to offer), but rather because we haven't had the time and resources to explore them in greater detail. Additional underutilized functionalities include provenance tracking, speech-to-text for transcription entry, and the ability to configure different vocabularies for separate segments of our corpus.

"Moving forward, the major question for SSDA as far as SH goes is where it fits into our workflow. There are a number of tools that are geared toward accomplishing some of what SH does - and likely better/more efficiently than SH - but none other that accomplishes all of them in a single interface. To some extent this is a result of the development that Jim has done with us in mind, but SH has always been an extremely ambitious tool that encompassed a wide range of functionalities. There are many projects that would benefit from using SH for any number of use

cases, and as far as I'm concerned the only issue is whether it is the best solution for any one dimension of this sort of work!"