Challenging Historical GIS through Unstructured Historical Primary Source Materials: Uncovering Sixteenth-Century St. Vincent, Aosta, Italy

James P. Schindling
Challenging Historical GIS through Unstructured Historical Primary Source Materials: Uncovering Sixteenth-Century St. Vincent, Aosta, Italy

James P. Schindling

Thesis submitted to the Eberly College of Arts and Sciences at West Virginia University in partial fulfillment of the requirements for the degree of Master of Arts in Geography

Approved by:

Trevor M. Harris, Ph.D., Chair
Kenneth C. Martis, Ph.D.
Matthew Vester, Ph.D.

Department of Geology and Geography

Morgantown, West Virginia
2016

Keywords: Historical GIS, St. Vincent, Geography, WVU
Copyright 2016 James Schindling
ABSTRACT

Challenging Historical GIS through Unstructured Historical Primary Source Materials:
Uncovering Sixteenth-Century St. Vincent, Aosta, Italy

James P. Schindling

The challenge to Historical GIS examined in this study is how to enable a system to accommodate sparse, unstructured, and spatially ambiguous data. The goal of this research is to take up this challenge by developing an Historical GIS that allows researchers to gain an understanding of a geography of the past. The study area for this research is the town of St. Vincent in what is now the Italian Alps and the time period is the sixteenth century. Three primary criteria were used in the selection of sixteenth-century St. Vincent as the case study for this research. The first was the availability of a rich, unstructured, and largely untapped set of data. These data are in the form of hand-written ledgers produced in the sixteenth century which include notarial records, birth records, and tax surveys. These ledgers record land holdings, property transactions, marriages, contractual agreements, and various disputes that took place among the inhabitants of the region. The second criteria was a geographical context that could be examined as a setting in which the events described in the source materials were acted out. This information is available in the form of maps of the region that were commissioned by the Italian government in the late nineteenth century. These maps identify towns, place names and land parcels, many of which are recognizable from the earlier texts. Finally, St. Vincent holds questions of interest to historical scholars related to social interaction, land ownership, and familial survival strategies. This work is a study in the incorporation of unstructured historical documents within GIS. Specifically, an investigation into the techniques that facilitate ingesting unformatted information into the highly structured requirements of current computer based information systems.
DEDICATION

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

I would like to express my sincere appreciation to my committee chair, Dr. Trevor M. Harris, for his advice, guidance, and patient review of my work. His experience, knowledge and demeanor were constant examples of true scholarship. Without his guidance this thesis would have been a far lesser work.

I would also like to thank my committee members, Dr. Matthew Vester and Dr. Kenneth Martis. Dr. Vester was kind enough offer his expertise on sixteenth-century St. Vincent and was patient enough to answer all questions that I brought to his office. Dr. Martis was kind enough to offer his support and his substantial credibility to my research.

Also, a thank you to all of the people I came to know at West Virginia University. My professors and fellow students were constant sources of support and encouragement throughout the process.

Finally, I would like to thank my family for their tolerance of my eccentric yearning to return to school.
TABLE OF CONTENTS

ABSTRACT .............................................................................................................................. ii
DEDICATION ........................................................................................................................... iii
ACKNOWLEDGEMENTS ...................................................................................................... iv
TABLE OF CONTENTS ......................................................................................................... v
LIST OF FIGURES .................................................................................................................. vii
CHAPTER 1 – Introduction .................................................................................................... 1
CHAPTER 2 – The Rise of Historical GIS and the Challenge of Historical Spatial Data .... 8
  Historical Geography ........................................................................................................ 8
  Historical GIS ................................................................................................................... 9
  Challenges of Unstructured Historical Data .................................................................... 12
CHAPTER 3 – Crafting a Spatially Enabled Database Solution ........................................ 18
  System Architecture Considerations ............................................................................. 18
  Processing the materials ................................................................................................. 19
    Map Processing ............................................................................................................. 20
    Text Processing ........................................................................................................... 25
  Exploration ..................................................................................................................... 36
  Mobile Access ................................................................................................................ 40
  System Components ....................................................................................................... 42
    Custom Coding ............................................................................................................ 44
CHAPTER 4 – The Case Study ............................................................................................. 45
  Selection of the Study Area ............................................................................................ 45
  St. Vincent ...................................................................................................................... 46
  Texts ............................................................................................................................... 49
  Maps ............................................................................................................................... 51
CHAPTER 5 – The System in Practice ................................................................................. 56
  Application of HGIS to sixteenth-century St. Vincent .................................................. 56
  Historical Materials ....................................................................................................... 57
    Space and Place .......................................................................................................... 57
    People .......................................................................................................................... 60
Inheritance .......................................................................................................................... 63
Land Use ................................................................................................................................. 64
Use of Additional Sources ...................................................................................................... 66
Additional Reporting Capability .......................................................................................... 67

CHAPTER 6 – Future Research Direction and Conclusion ...................................................... 69
Future Directions .................................................................................................................. 69
  Functional Enhancements .................................................................................................. 69
  Enhanced Support for Historical Geography and Historical Research ......................... 70
  Technology Enhancements ............................................................................................... 71
Conclusion ............................................................................................................................. 71

REFERENCES ......................................................................................................................... 73

APPENDIX A .......................................................................................................................... 76
LIST OF FIGURES

Figure 1: Example of sixteenth-century hand-written notarial record........................................14
Figure 2: System Architecture........................................................................................................18
Figure 3: Data Ingest Pathways ......................................................................................................19
Figure 4: Map Six ..........................................................................................................................20
Figure 5: Control Point Selection ..................................................................................................21
Figure 6: Aerial imagery of modern terrain....................................................................................22
Figure 7: Detail of map 6, Vector conversion...................................................................................23
Figure 8: Detail of map 6, Creating polygons..................................................................................24
Figure 9: Primary forms used for information management and exploration ..................................25
Figure 10: Research Process .........................................................................................................27
Figure 11: Menu option on the transcription listing used to create new entries.............................28
Figure 12: Form used to manage individual transcription...............................................................28
Figure 13: An example of the document image tab......................................................................29
Figure 14: An example of a transcription entered into the system..................................................30
Figure 15: An example of the Translation interface.......................................................................31
Figure 16: Form used to maintain information about people..........................................................32
Figure 17: Results of scanning the text for people’s names.............................................................32
Figure 18: Excerpt from the thesaurus used match names with multiple spelling variants...........34
Figure 19: The property editing interface.......................................................................................35
Figure 20: Mapping form...............................................................................................................36
Figure 21: Example map layers extracted from the source maps.....................................................37
Figure 22: Example of a query locating vineyards with a medium confidence level.......................37
Figure 23: Example of a spatial query.............................................................................................38
Figure 24: Spatial query options from the people form..................................................................39
Figure 25: Spatial query options from the transcription directory form.........................................39
Figure 26: Mobile access screen....................................................................................................40
Figure 27: Browser interface.........................................................................................................41
Figure 28: Service architecture components..................................................................................42
Figure 29: Summary data model for the Transcription database....................................................43
Figure 30: Map Six details..............................................................................................................52
Figure 31: Map Six signatures........................................................................................................53
Figure 32: Mapped Structures........................................................................................................54
CHAPTER 1 – Introduction

Over time memories fade, relationships fragment, and once vital records collect dust in obscurity. The diligent and detailed recordings of births, christenings, marriages, property holdings and burials are set aside, and with them the knowledge of generations that they document. Even the language and handwriting of each era change almost as if they have made a conscious effort to obscure what was previously recorded. But evidence persists; the heavy stone walls and field systems that helped support an agrarian way of life may fall or be appropriated for other purposes, but traces remain. Yet documents that record these places can be resurrected and interpreted by a new generation of researchers. Maps created for different purposes and at different times can be analyzed to provide links to prior human geographies.

For most social science researchers, neatly structured and spatially precise tabular data offers an ideal information source. Rows and columns of well labeled and consistently measured values are ready made for analysis. However, history is not always so accommodating in providing information in this way. Much of history is 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, and even the use of obsolete units of measurement.

Applying Historical GIS (HGIS) to piece together these disparate clues and to provide a glimpse into the past is challenging. As Knowles and Hillier (2008, p. 7) point out, “Geographical questions drive a significant part of historical inquiry”. Arguably, it is through the use of Historical GIS that the spatial information embedded in these unstructured data can be
drawn out and given geographical form. Even though the historical materials are unstructured, they contain valuable information such as place names, personal names, measurements and actions that can be parsed and ingested into a GIS database. However, while the authors of these historical data knew through lived experience the locations and significance of the places that were referenced in the documents, current researchers do not have such access to the intimate understandings of the places or actors of the time. By applying Historical GIS it may be possible to visualize and analyze not only the geographic locations recorded in the documents, but also develop a deeper sense of the significance these geographies.

History is rich in spatial information, yet traditional historical research often glosses over the geographical significance of this information. At best, events may be shown as points on maps, however, it is often the case that events are simply said to have occurred at some seemingly disembodied location. One approach to provide a geographical context and facilitate spatial enquiry to historical events is to supplement the printed narrative with maps. An excellent example of this is the recent publication of an annotated version of Herodotus’ Histories that includes 137 maps showing the locations of significant places and events described in the original text (Strassler, 2009). Geographical Information Systems (GIS) offer the ability to enhance the association between events and places by facilitating the creation of maps and offering a repository in which to store related information. It has been said that GIS should be regarded more as a database technology than a mapping technology (Gregory and Healey, 2007). However, this focus on technology and map production has also led to many criticisms of GIS and its applicability to the study of the humanities, where it has been argued that the traditional positivist epistemology that underlies the technology is fundamentally at odds with the humanistic approaches pursued in historical research (Harris et al., 2010). Recently, however,
this perceived divide between the qualitative critical approach to humanities and the quantitative analytical approach of GIS has come to be questioned (Bodenhamer et al. 2008, 2015). Rather than being antagonistic and in competition with each other for legitimacy, awareness is growing that these approaches can in fact be complementary to produce a comprehensive understanding of human activity – past and present. In this vein, GIS is increasingly seen in the humanities as a mechanism to support historical research (Knowles and 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, 2008; Goodchild, 2008; Gregory and Healey, 2007). Beyond being a mapping technology, Historical 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.

Gaining a greater sense of historical events means not only building a description of the setting and constructing maps, but also developing an understanding of the contemporary geographic and environmental conditions that may have influenced the actors. Examples of this type of research can be found in the Historical GIS work of Knowles (2008) and Harris (2002). Knowles built a virtual reconstruction of the Civil War battlefield at Gettysburg and used viewshed analysis to help explain General Lee’s seemingly uncharacteristic order to have General Pickett charge directly into the strength of the Union army. Harris (2002) used similar techniques to investigate what appeared to be the random placement of iron-age habitations and farmsteads in relation to the location of the Danebury hillfort in Hampshire, England. In these examples, the terrain and land cover characteristics themselves were important information sources.
Other researchers have also developed deeper understandings of historical events through the use of structured historical data sources. An example of these types of data is the *National Historical Geographic Information System* (University of Minnesota, 2010). This site provides structured historical census and demographic data, along with GIS compatible boundary files. The use of these structured source materials aligns neatly with the database structures that underlie GIS technology. Fixed boundaries and tabulated numerical values are the hallmarks of traditional Historical GIS work. However, much historical data is far less structured and considerably more ambiguous in its recording of space and time. 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 et al., 2015). For example, in the case of personal names, individuals may be referred to in the primary source documents by their relationship to others rather than by their own full names. Equally, places may be given names based on their geographic proximity to other locations or based on ownership. Disambiguating the relative and spatial references in historical text becomes one of the significant challenges to the use of these data within Historical GIS (Smith and Crane, 2001). These types of unstructured data demand that the database structure, analysis, and visual representation accommodate these discrepancies.

The use of modern maps that are georeferenced and based on standard coordinate systems often provides an additional data source for Historical GIS research. These maps may be
in analog form which can be scanned or may be available in a digital format that can be ingested directly into a GIS. However, historical maps were invariably created without any coordinate systems and therefore provide additional challenges to GIS because they cannot be directly imported or re-projected without considerable pre-processing (Bodenhamer, 2008). Most historical maps must first be scanned into digital form and line work that is missing or has been corrupted must be corrected. In addition, sets of maps might need to be digitally stitched together into a mosaic to capture the entire spatial context (Lafreniere and Rivet, 2012). Because of this, several pre-processing steps must be undertaken to prepare historical maps for use in a GIS.

The challenge to Historical GIS examined in this study is how to enable a system to accommodate sparse, unstructured, and spatially ambiguous data. The goal of this research is to take up this challenge by developing an Historical GIS that allows researchers to gain an understanding of a geography of the past. The study area for this research is the town of St. Vincent in what is now the Italian Alps and the time period is the sixteenth century. Three primary criteria were used in the selection of sixteenth-century St. Vincent as the case study for this research. The first was the availability of a rich, unstructured, and largely untapped set of data. These data are in the form of hand-written ledgers produced in the sixteenth century which include notarial records, birth records, and tax surveys. These ledgers record land holdings, property transactions, marriages, contractual agreements, and various disputes that took place among the inhabitants of the region. The second criteria was a geographical context that could be examined as a setting in which the events described in the source materials were acted out. This information is available in the form of maps of the region that were commissioned by the Italian government in the late nineteenth century. These maps identify towns, place names and land parcels, many of which are recognizable from the earlier texts. Finally, St. Vincent holds
questions of interest to historical scholars related to social interaction, land ownership, and familial survival strategies.

GIS offers great advantages for the study of history given the inherently spatial character of human activity. However, the unstructured nature of historical materials, such as those from sixteenth-century St. Vincent, sits uneasily with the structure required by the computer systems on which GIS technology is based. This research challenges Historical GIS to move beyond the requirement for highly structured spatial datasets toward the use of unstructured historical sources typical of early-modern societies. This research explores the capabilities of Historical GIS as a means of comprehending the spaces and places of times gone by. The Historical GIS developed for this project will be used to tease out a glimpse of a bygone society and its geography by locating the actors, places and events recorded in unstructured materials. As Ian Gregory (2014) notes, “A key test of the effectiveness of historical [GIS] research is that the work that it produces should be of interest not only to HGIS specialists but also to an audience of subject specialists…”. The challenge for this study is to ingest information from sparse and unstructured sources into an inherently quantitative environment while retaining and exposing the essence and historical value of the records.

The research questions guiding this research are designed to lead the project toward the development of an Historical GIS that goes beyond the normal application of off-the-shelf GIS tools. Specifically, this study will address the following questions:

1. What is Historical GIS and how has it contributed to understanding the geographies of the past?
2. What are the issues involved in studying the historical geography of early-modern places and what are the challenges to be confronted in using Historical GIS to examine primary historical source materials?

3. What are the available primary source materials for sixteenth-century St. Vincent and how might they be ingested into a Historical GIS?

4. How might Historical GIS be used to recreate the cultural landscape, landholding strategies, and familial relations of sixteenth-century St. Vincent?

5. What insights does the application of Historical GIS offer for the historical geography of St. Vincent in the sixteenth century, including furthering the understanding of the cultural landscape, land ownership, familial power relationships and inheritance strategies?

6. How valuable and appropriate was HGIS in the analysis of the historical geography of sixteenth-century St. Vincent?

To address these questions, this paper will proceed by first discussing the evolution of traditional Historical GIS and identifying the approach taken by this study that makes it unique. This is followed by a discussion of the technical design of a system that supports the management and exploration of unstructured historical materials. Next, the case study of St. Vincent is discussed, indicating the key challenges that it offers. Finally, the application of the system to St. Vincent is examined, highlighting the ways in which the technical design supports exploratory historical geographical research.
CHAPTER 2 – The Rise of Historical GIS and the Challenge of Historical Spatial Data

This work is a study in the incorporation of unstructured historical documents within GIS. Specifically, an investigation into the techniques that facilitate ingesting unformatted information into the highly structured requirements of current computer based information systems. To understand the contribution of this research to Historical GIS, it must be situated in the broader context of Historical Geography.

Historical Geography

Historical Geography has a long pedigree that arguably goes back to the writings of Herodotus. His works were as much descriptions of peoples and places as they were recordings of events. The spatial context and geographical conditions of his subjects were prominent in his descriptions of their actions. Butlin defines Historical Geography as;

“the study of the geographies of past times, involving the imaginative reconstruction of a wide range of phenomena and processes central to our geographical understanding of the dynamism of human affairs, such as change in the evaluation and uses of human and natural resources, in the form and functions of human settlements and build environments, in the advances in the amount and forms of geographical knowledge, and in the exercising of power and control over territories and peoples.” (Butlin, 1993)

In its modern form, Historical Geography has its roots as a branch of history that filled in the geographical and environmental backdrop to the chronological documentation that was customary in historical writing (ibid.). The place of Historical Geography in scholarship, however, remained open for debate. Its function has been seen as ranging from a study of the effect of geographical conditions on historical events to the study of historical events of
geographical facts. Regardless of the debate within the geographical community, historians have traditionally found the discussion somewhat irrelevant to their work and, in some ways, distracting from their essence of the research.

Over time it became recognized that situating history within a spatial context added a new dimension to discussion and debate of historical materials. The idea that space and geography were themselves factors in historical events led to the investigation of new sources of evidence for historians including maps, relict features on the landscape, and imagery. Maps themselves came under new scrutiny as artifacts from which the unspoken message of their authors that could be deconstructed by examination of their content and structure (Federzoni, 2013). Maps began to be studied for their codes of meaning, technical production, and cultural production and they themselves thus became historical artifacts (Drucker, 2011).

**Historical GIS**

A new investigative capability became available for historical geographers in the 1970’s with the introduction of computers (Butlin, 1993, p. 88; Bailey and Schick, 2009). Large volumes of digital information could be summarized and analyzed to produce what was seen as either simplistic positivist results or new avenues of research depending on the inclination of the researcher. Tabular data such as census reports could be quickly processed to generate population statistics, racial statistics, religious statistics; statistics on anything for which there was available data. This capability offered new ways of looking at historical information in aggregate to help draw out patterns of actions and events.
Where Historical Geography has a long history, Historical GIS is in relative infancy. The existence of Historical GIS is predicated on broad accessibility to computing power and on 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 it can be exploited for purposes of historical research continue to arise (Knowles A, 2005, Schlichting, 2008;). Initiatives such as the Stanford Spatial History Project have appeared as incubators for creating new ways of looking at, and “doing”, history (White, 2010), seeking to examine historical events in ways that augment traditional historical research.

All historical events occurred somewhere and thus they have a location component as part of their story. Historical GIS offers the advantage of incorporating the spatial component explicitly within an analysis allowing the patterns and distribution of events to be addressed directly (Gregory and Healey, 2007, p. 644). Using Historical GIS it is possible to combine numerical information, such as census data, and geographical information to produce a variety of location-based visual representations. A familiar type of representation is the choropleth map which is a visually straightforward way of representing color coded numerical data in a spatial format. Other representations, such as the cartogram, were developed which sacrifice locational accuracy while emphasizing the data by scaling the visual representation in proportion to the numerical attribute value.

An additional benefit of current GIS technology is the ability to visualize events in a pseudo three dimensional (2.5D) setting. For historical research this can offer a great deal of additional insight. For example, the technique of draping the locations of the opposing armies on a 2.5D rendering of the landscape was used to produce a visualization of the opponents at the Battle of Hastings (Gregory and Ell, 2007, p. 111). This example also provides a case study for
the need to integrate spatial analysis using GIS with traditional historical research. Examining the GIS rendering alone indicates that the tactical advantage based on topography was held by the English, implying better tactical planning and a battlefield advantage. The result of the battle, however, was a victory for the Normans under Duke William, demanding a closer look at the documentary evidence to explain the outcome.

The use of GIS for historical research, or any research, is based on the availability of data that lend themselves to computer based processing; meaning that the data are tabular, quantifiable, complete and consistent. GIS data are actually made up of two components; the most obvious is geographical information (GI), or the information that specifies a location on the earth’s surface. The second component is information about the location, referred to as attribute data. Attribute data identifies relevant characteristics about the location such as the name, population, or agricultural production capacity. In a GIS, this attribute information is represented as rows 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 current GIS (Gregory and Ell, 2007).

The production of maps of all styles has been made far easier with the availability of web accessible geographical information layers such as topographic layers, administrative boundary layers, and road layers. Along with these layers there are highly accurate elevation data that can be used to produce 2.5D landscape renderings. These elevation data are publicly available from both 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 in a way that makes it far more accessible than what could have been imagined only a decade ago.
**Challenges of Unstructured Historical Data**

Producing maps using a GIS and representing structured spatial data have become relatively straightforward and common activities. The challenge for this research, however, is to develop an Historical GIS using highly unstructured texts and hand-drawn maps that were produced at a time when computer systems had not been considered and when data was stored in the form of narrative rather than spreadsheets.

**Historical Maps**

Relatively sophisticated map making was taking place in Europe during the sixteenth century, even if by modern cartographic standards it may appear somewhat crude (Federzoni, 2013). Since map making required a high degree of skill and was very much a manual process, the cost of producing maps limited their commission primarily to wealthy landholders, nobility, and the church. The result is that the focus of most maps from the sixteenth century was on land holdings and structures owned by the elite. The fields and homes of the common workers were often given only cursory consideration and were only included if they in some way contributed to the narrative that the patron wanted to convey. In this sense, maps from the sixteenth century are in many ways as useful as pieces of ‘historical text’ as they are as sources of cartographic information. They provide insight into the types of features that were viewed as significant, either monetarily or as social status symbols. Examples of many of these historical maps can now be accessed directly from the web and can be downloaded as high resolution raster images at no cost (Cartography Associates, 2016).

The process of using historical maps, which on the surface may seem straightforward, can involve significant effort even with the use of the most sophisticated software currently available (Bitelli et al., 2009). In the case of hand-drawn historical maps there are added
complexities. The maps are invariably produced without a standard system of spatial reference and for practical reasons large areas are often drawn as several separate maps. In situations such as this each map must be georeferenced individually and then combined into a single mosaic. An example of this process is given by Lafreniere and Rivet (2012) in the work that they performed using historical maps of Sandwich, Ontario, Canada from the early nineteenth century. In their work they were able to demonstrate a methodology for assembling a set of hand drawn maps into a single consolidated map. An in-depth evaluation of the cartographic considerations involved in this process can be found in a work by Bitelli et al., (2009). The Bitelli paper discusses issues surrounding the use of different coordinate systems and cartographic projections and how they can introduce potential errors into mapping the process. The georeferencing of historical maps can be extended to the reconstruction of 2.5D landscapes (Brumana and Achille, 2007). For historical and archaeological research this approach offers a new avenue for modern scholars to gain greater insight into the area of study (Knowles and Hillier, 2008).

Unstructured Texts

Up to this point the discussion has been on the use of tabular data sources to provide location and attribute information with in a GIS. There is, however, another class of data source from which these types of information can be extracted and that is of unstructured text and data. With the recent advent of ‘big data’ in the form of Twitter feeds, Facebook posts and other social media sources, a new field of research has developed around extracting information from unstructured sources. There are a variety of techniques that have been employed and also a variety of motivations ranging from disaster response to market analysis. In all cases the
objective is to first recognize the pieces of information that are of importance and to then
reorganize them into a format that is conducive to further analysis. Because of the variability
involved, recognizing the key information is where a large portion of current research into ‘big
data’ analysis is targeted. The complexity of working with these unstructured sources lies in the
fact that there are no formal rules regarding content, structure, language, syntax, or spelling.
There is also no moderator to control the quality of the information; anyone can post anything
with little regard to validity (Chen, 2015).

There has recently been research into automated information extraction software within
the humanities (Giusti et al., 2007; Smith and Crane, 2001). In many cases these systems rely on
the availability of large textual corpuses that can be parsed and analyzed based on known rules of
syntax, semantics, and punctuation. These corpora allow systems to be trained and validated to
the point where they can, with relative consistency, identify the names and locations of events
and places of importance. The Giusti (2007) paper offers an interesting approach in that it does
not attempt to discern meaning from individual words but rather converts letters and letter
combinations into standard phonetic notation that is used for matching terms. This approach has the advantage that phonetic
rules can be defined and the results tested regardless of language.

Extracting information from hand-written historical texts
has been demonstrated using tabular data such as census records
(Fu et al., 2014). Information extraction becomes more complex
with older records because they are invariably hand written with
data embedded in descriptive prose, making it difficult to simply find location references or

Figure 1: Example of sixteenth-century
hand-written notarial record.
names of individuals (Figure 1). In addition, the large volumes of data that are required to train information extraction algorithms simply do not exist from earlier times. In areas like the Alps in early-modern times many of these documents were written in dialects that have changed or disappeared over time. Many were also written in a paleographic style that can now only be interpreted by a small number of individuals. These issues of transcription and translation create additional layers of uncertainty when trying to understand the content of the materials themselves and make their use within a GIS more difficult.

**Toponyms**

The problem of using historical texts in a GIS is further complicated by the use of locally know toponyms rather than formally defined place names. Over time these local toponyms often fade from use and are replaced by others that become more relevant. For example, an event may have been recorded as having occurred near a communal bread oven. Generations later that oven may no longer exist and its place on the landscape forgotten. The search to identify toponyms within historical texts takes on greater significance in the context of historical GIS because toponyms are often the only means of extracting location information.

The use of toponyms and, more broadly, ways of referencing places has developed in various ways even within the same geographical area and time periods. Toponyms attached to the landscape make up the cartographic lexicon of the inhabitants and reflect the cognitive framework within which they lived their daily lives. Toponyms were often “folk classifications” recorded in the vernacular of the inhabitants. Smail (1999) presents an in-depth study of the evolution of naming systems in sixteenth-century Marseille. Smail presents the idea of different mental and social templates that different groups unconsciously apply when referring to places. Smail suggests that aspects of a given place that are of most importance to an individual’s daily
life and world view become incorporated into the way that individual defined and described his/her surroundings. For a merchant, for example, an area may be defined in terms of a market and its constituent trading booths, each specializing in a particular type of good, while for a craftsman the location of individual shops may be the basis for understanding a local geography. For members of the clergy these local activities may be overshadowed in importance by the various parish divisions of the region leading to toponyms based on parish affiliation rather than function. This overlapping mix of social, religious, and administrative areas can be seen in the naming conventions used in the records as the notaries adapted their records to suit the mental and spatial templates of involved parties.

The development in the Alps of named streets with numbered addresses came about as an unconscious evolution based on the need for notaries to develop a common nomenclature that could be applied consistently regardless of the background of the individual parties involved in an agreement (Smail, 1999). The use of numbered addresses also made it possible to access and understand locational meaning by anyone who had an interest in tracing property ownership without relying on an understanding of the template through which the original parties knew the property.

Many early modern documents, such as those from St. Vincent, offer only local toponyms, or *lieudit*, as clues to actual geographic locations. Some of these names have persisted into the present day as formal place names while others such as ‘*Loz Champ de la Cuisine de Sainct Mauris*’ have faded from memory along with the features that they once referenced. Also, many of the more formally named places from the documents appear to have been recorded with various spellings, based possibly on the preferences of the recording notaries or some other factors that are now unknown.
The challenge for this research is to develop an Historical GIS based on highly unstructured texts and hand drawn maps that were produced at a time when computer systems had not been considered; a time when data was stored in the form of narrative rather than spreadsheets and places were identified based on a lived understanding of the region. The ability to work with these types of data within a comprehensive, integrated system is a capability that is not inherent to off-the-shelf GIS software and has not been investigated as part of traditional Historical GIS research. The following chapter describes a software infrastructure and user interface that addresses the use of these unstructured data within a unique spatially enabled exploratory information system.
CHAPTER 3 – Crafting a Spatially Enabled Database Solution

System Architecture Considerations

Historical materials such as those from sixteenth-century St. Vincent pose unique technical challenges that require processing and preparation prior to use within a GIS. Making the resulting information accessible to historical researchers requires the development of an information management system that supports ingesting map data, textual materials, and mobile uploads of ancillary information. Once loaded, these data must be made available for use in a fully integrated exploratory research tool that supports a variety of information access paths. Figure 2 presents the architecture for the

Figure 2: System Architecture. Functional requirements include map processing, text processing, information exploration, and mobile collaboration.
system developed this study to provide this capability. The primary functional areas include map processing, text processing, information exploration, and mobile collaboration; all centered on an integrated data server.

**Processing the materials**

There are three separate paths by which data can be ingested into the system; one for loading maps, one for texts, and one for collaborative contribution of materials. Each of these pathways has its own set of required tasks as shown in Figure 3. The map loading process is a onetime activity that is foundational to the function of the system while text processing and collaborative contribution are ongoing activities that add new information as it comes available. The following sections discuss these paths and how the system has been developed address each one.

*Figure 3: Data Ingest Pathways. The tasks in parenthesis are optional depending on the specific data. *Scanning was performed by the archive curators.*
Map Processing

Figure 4 shows an example of a map for part of St. Vincent that 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 generated from scans of the original hand-drawn maps. All of the map processing discussed below was performed using the ArcMap software package from ESRI. This package has a variety of tools that support raster image processing, raster to vector conversion, and map manipulation.

Figure 4: Map six showing the date that it was recorded, the scale, and an indication of the bordering maps.
**Loading and Cleaning**

Loading the scanned maps required copying them to a folder on a disk drive accessible to the ArcMap software. It is apparent from visual inspection of the ingested maps that there are a lot of markings on the scanned images that do not represent actual map features. These include various written labels, notations, and smudges that are artifacts arising from use and age but were captured by the scanning process. To make the maps usable within a GIS the first task was to clean the unwanted markings from the images. This was a manual process that involved using several raster processing tools within the ArcMap software package.

**Georeferencing**

Because these maps were hand-drawn and were created prior to the development of standard coordinate systems they must be adjusted so that they match the actual geography of St. Vincent. This process, called georeferencing, was performed by matching 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 specialized GIS functions to stretch and warp the original map so that it aligns with the actual coordinate geography (Figure 5). This process requires sufficient control points such that the map is modified to the correct proportion. In many cases there were existing features such as buildings or roads that offer
accurate reference points. However, many maps are of areas that are now covered with vegetation and trees that obscure surface features. This can be seen in Figure 6 where a scan of map eight is overlaid on current satellite imagery. In this example there are many clearly identifiable features existing 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. When situations like this were encountered, control points were 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.

The result of the georeferencing process is 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 general locations of villages this was not too surprising, however field boundaries and traces of the canal system also align to a very large degree providing a level of confidence in the georeferencing results.
**Convert to Polygons**

With the raster map images georectified, the next step was to construct vector lines that correspond to individual parcel boundaries and other features. This conversion is shown in Figure 7. Notice that most of the unwanted markings in the raster images have been successfully removed and the boundaries are clearly delineated. The lines resulting from this process were then used as the basis for creating polygons for each of the features. These polygons are the final representation of the spatial extent and location of each of features that were drawn on the original maps.

*Figure 7: Detail of map 6 showing (a) the raster version and (b) the resulting vector lines*
Assign Feature Details

Each resulting polygon represents the geographical information for a particular feature from a map (Figure 8a). In addition to this spatial information, the maps contain identifying numbers for many of the features and visual clues as to what the features are. On the maps of St. Vincent there are 17 discernable types of features, four of which can be seen in Figure 8b. In this figure the green polygons identify parcels of land, the brown indicate structures, the gray indicate roads, and the blue indicate canals. The maps also contained the names of roads, waterways, towns, and some churches which had to be manually entered into the attribute data. It total, over 25,000 polygons were generated from the 67 hand-drawn maps. Of these, over 21,000 were numbered on the source maps, each requiring manual entry of the number values.

Merge Maps

There were 67 scanned maps produced which in combination constituted the entire nineteenth-century survey of the region. Thus each of the processing steps mentioned above had to be repeated for each map resulting in 67 individual maps (or feature classes) within the GIS requiring an average of approximately two and one half hours per map. The resulting maps were then consolidated into a single comprehensive map and mismatched boundaries were adjusted manually. In addition, features such as roads, that crossed from one map to another were merged.
For most roads, canals, and creeks the merge process had to be repeated across several maps. With these tasks complete, the comprehensive set of geographical information was available to serve as the backdrop for the events that were recorded in the texts.

**Text Processing**

There are two driving considerations behind the design of the system interface. The first consideration was that the interface must support the process of capturing and managing the information contained in the historical texts. The second consideration was that it must facilitate and even encourage information exploration. To address these requirements the user interface is designed based around four primary forms: the main *Transcription Directory form*, the *Transcription Detail form*, the *People form*, and the *Mapping form*. These forms, shown in Figure 9, are used in combination to perform all the tasks required to enter, maintain, and explore the transcriptions. The *Transcription Directory* (Figure 9a.) is the primary form for navigating the system and contains a listing of document ledgers and the individual transcriptions they contain. The lower left section of the form displays possible toponym and territory matches for
selected transcriptions. The lower right section holds the names of individuals, territories, and
toponyms associated with the selected transcription, as well as a listing of related property
information and images of the source document pages.

The *Transcription Detail form* (Figure 9b.) is the main interface for entering and editing
individual transcriptions. The left section contains metadata that is used to identify the
transaction, detail extracted from the raw transcription, and other relevant notes related to the
transcription. The right section of the form holds images of the source document pages, the
transcribed text, and translated text. The *Person form* (Figure 9c.) is used to maintain the list of
individuals identified in the source materials. Names are arranged in both paternal and maternal
family hierarchies. This form is also a starting point for examining family property holdings and
for identifying related transcriptions. The *Mapping form* (Figure 9d.) provides a visual interface
for presenting transaction related data in a spatial context. The *Mapping form* also contains nine
additional map layers that can be turned on or off by the researcher to support exploratory
interaction. The left section of the form contains a hierarchical listing of the territories and
toponyms from the original map ledgers along with options to query the map either by specifying
an area of interest or by land use type. This form and its function in data exploration is discussed
in more detail in the *Exploration* section below.

**Entering and Managing Transcriptions**

While there is no workflow imposed by the system for adding data, the basic set of tasks
and their informal sequence is shown in Figure 10. It is important to note that these tasks can be
performed either synchronously or asynchronously. That is, an individual document can be taken
through these steps from start to finish prior to moving on to the next, or multiple records can be
processed through a single step then taken as a group through subsequent steps. For example,
there may be a member of the research team who specializes in document transcription and another who specialized in translations. In this case the system would allow the transcriber to perform the first three steps while the translator takes over to perform the translation. Other team members can then perform the remaining tasks at their leisure. The system architecture also allows members who are geographically dispersed to work remotely on the same set of texts, and thereby facilitating collaboration between a larger group of researchers.

The entire process is facilitated by allowing a user to navigate seamlessly between all facets of document management, data exploration, and information extraction. The following sections demonstrate how these tasks are implemented within the interface.

Figure 10: Research Process. The steps in this process can be performed either synchronously or asynchronously and can involve as many researchers as is appropriate based on the texts and available skill sets.
Step 1: Creating a new record

The first step in processing a text is to create a new blank record in the system in which to hold the details. This step is done by selecting the ‘Add New Record’ menu option on the Transcription Directory form (Figure 11) which opens the Transcription Detail form and allows the researcher to begin entering information.

The Transcription Detail form is the primary interface used to maintain the details of individual transcript entries. The form allows the researcher to record details that can subsequently be used to categorize, summarize, and search the registry entries. Figure 12 shows the blank form as it appears when a new transcription entry is first created. Much of the work performed in subsequent steps takes place using this form.
Step 2: Importing document images

An important piece of information that can be stored within the system is the imagery of the corresponding register pages. Although it is not required, document images provide a very useful reference source for the transcription process and for subsequent historical reference. The *Transcription Images* tab of the edit form (Figure 13) enables the user to import imagery of source documents. This interface also has options that allow the user to zoom and pan around the images and to open the images in a separate window to support the transcription process. Because many transaction records span several hand-written pages the interface allows multiple images to be loaded per transcription which the user can easily move between.
Steps 3 and 4: Transcribing and Translating Text

Once the transcription images have been imported the researcher can begin the task of transcribing the text (Figure 14). This is very much a manual activity that, in the case of the St. Vincent documents, requires an understanding of the paleography and local patois of the time. While it is possible to perform a verbatim transcription of the script with only an understanding of the paleography there are cases where the text is obscured or missing. In these instances an understanding of the patois, Valdôtain, can potentially allow the researcher to reconstruct the missing information based on context and knowledge of the surrounding text.

A preliminary automated translation can be performed from this tab as well. Because sixteenth-century Valdôtain is somewhat similar to modern French a direct translation of the text from French to English produces results that contain some English words, however the majority of the document, and certainly the meaning, remains untranslated. To support a greater degree of success in translating the text, the automated process built into the system pre-translates the text using a built-in Valdôtain to French dictionary. This dictionary can be updated by researchers with new words and spelling variants as they are encountered.

The Translation tab in the detail interface is where the researcher finalizes and maintains the translated text. The translation can be entered from scratch or can be based on modifications...
to the results of the automated translation. This tab, as well as the *Transcription* tab, also support
the entry of notes. This feature can be useful for recording the choices made by a researcher who
performed the work or for recording other information that may be pertinent to understanding the
text.

**Step 5: Extract Person Information**

Once a text has been transcribed or translated, the system can be directed to scan the
text for known people or places. The repository of
matching person names is dynamically populated
as new individuals are encountered. In this way, the
system heuristically ‘learns’ how to identify names
in the texts as more transcriptions are processed.
The result of this functionality can be seen in
Figure 15. In this example, a search was performed
for person names in the transcription. The
highlighted words in the text itself indicate matches
with known surnames or given names. The
highlighted names can be further explored by the researcher to determine if they refer to a
previously identified individual or if they should be added to the names repository. The grid in
the *Text Search Results* window shows a list of possible matches that already exist in the
repository. This list is built by scanning the text for given name / surname combinations that
occur in close proximity to one another. In this case the given name “Jean Louis” and the
surname “Provaney” are found together in the text and are a possible match for any of several “Jean Louis Provaney” in the repository.

The system also supports the recording of various details regarding individuals encountered in the historical texts. The primary items of interest are usually gender, family relationships, affinal relationships, birth dates, and death dates. As more documents are processed, more of these details begin to accumulate and new individuals are encountered. Maintenance of the directory of people and their related details takes place through the use of the People form (Figure 16). This form is divided into two main sections: the left side of the form lists all of the people by both paternal and maternal lineage. The right section of the screen displays the details of the currently selected individual and allows updates to be made. In addition to adding and maintaining information about individuals, the form also allows the user to

Figure 16: Form used to maintain information about people encountered in the texts. Note the menu that allows the user to show related transcriptions and property locations.

Figure 17: Results of scanning the text for people’s names. The highlighted words in the transcription text are names. The names listed in the grid are matches from the repository.
navigate to a list of transcriptions related to the selected individual and display the locations of property holdings on the map interface. As individuals are added to the directory they can be linked to transcriptions by simply dragging them from the listing and dropping them on the transcriptions Associated People tab. This linkage is particularly useful during research to support the exploration of family holdings and their spatial distribution.

**Step 6: Extract Place Information**

Extracting place information involves linking toponyms, or lieudit, from the sixteenth-century documents to toponym and territory names listed in the nineteenth-century map indexes. With this association made, the properties can be shown in the map interface and spatial relationships can be determined. The process of linking transactions to places begins by locating place names in either the transcribed text or in the translated text and searching the map index for matching toponyms or territories. Because of the large number of names involved, it is not always feasible to perform this activity manually. To facilitate this search process the system employs two distinct search mechanisms; freetext matching and pattern matching.

Freetext matching is based on thesaurus technology built into the underlying query engine. The technique involves developing a thesaurus of place name synonyms and directing the system to use it when performing queries. This mechanism is useful for performing fuzzy searches based on toponyms encountered in the source texts. Figure 18 shows an example of a portion of the thesaurus that directs the system to match on eight variations of the name “Champbiller” and three variations of “Champagne”. By convention, the names from the map indexes are used as the ‘correct’ names during matching operations because unique physical map locations are associated with unique names in the indexes whereas the notarial records often
contain a variety of spellings when referencing a single location. Thus, a query on either the name “Champmilliet” or “Champeille” from a notarial record would match with the standard toponym “Champbiller” from the map index. This search and match functionality is used both when scanning the transcriptions for names and when querying the toponym table for matches.

There are many cases where new spellings are encountered in the texts that have not been previously entered into the thesaurus. When this occurs the researcher must try to match names based on a subset of characters or possibly on spelling variations in which letters are substituted based on phonetic similarity. The Pattern matching technique allows the researcher to enter letter sequences and optional wildcard characters that are then compared with names from the map indexes. For example, the name “Excontre” from a text does not find a match. In this case the researcher could perform a search on the pattern “e%tre”, where the “%” acts as a wild card, resulting in matches on the toponym “Ecantre”.

Once a territory or toponym has been identified it can be linked to transcriptions by simply dragging it from the listing and dropping it on the transcriptions Territory or Toponym tab. This linkage is used subsequently to support the spatial exploration of the historical texts.
Step 7: Identify Property Types

Recording the types of properties involved in transactions is performed on the main Transcription Directory form. When property information is encountered in the texts the researcher can associate it with the currently selected transcription by going to the Property tab and selecting the Add option from the context menu (Figure 19a). When the Add option is selected a form appears that allows the researcher to enter the type and quantity of property along with any notes that might be applicable. There are eight units of measure to select from when associating property with a transaction. The default is ‘Unit’ since many of the transcriptions indicated a property type but not a size. Since the units of measure used in the notarial records are not consistent and are different from the metric system used today, the system is designed with a unit-of-measure conversion capability. When fully implemented, this conversion feature will allow researchers to look at property transactions in consistent and familiar terms.
**Exploration**

Data loading and management are large components of the systems design. However, the raison d’être for this project lies in its ability to enhance historical research by providing a spatial context for the events recorded in the unstructured source materials. The interface used to provide this capability is the mapping form shown in Figure 20. At its simplest, this form allows the user to view the different feature types that were constructed as part of the map processing. The form also allows the researcher to zoom, pan, and rotate the display in three dimensions. Figure 21 shows an example of a section of map data in the vicinity of St. Vincent, both with and without a satellite image background. In this example, it is possible to see the locations of land parcels, structures, canals, a river, and roads as they existed on the nineteenth-century maps. The layer selection list in the upper right corner of the display can be used by a researcher to turn on or off individual layers and to re-sequence their display order based on specific requirements.

*Figure 20: Mapping form. The left section lists the territories and toponyms extracted from map indexes. The right portion is an interactive 3D map interface.*
Two different query options are also built into this form: one allows the researcher to identify locations where specific types of property are mentioned and the second allows the researcher to perform a spatial query based on drawing a polygon on the land. An example of a query based on property type is shown in Figure 22. To perform this query, the researcher first selects a type of property from the dropdown list in the upper left portion of the form and optionally checks the box to show related transcriptions. Clicking the Query button runs the query and highlights locations of the matching territories and toponyms. This query feature is useful for identifying potential land use patterns implicit from the spatial distribution of the various property types.

Figure 21: Example map layers extracted from the source maps. a) without satellite imagery, b) with satellite imagery.

Figure 22: Example of a query locating vineyards with a medium confidence level.
The spatial query option (Figure 23) is run by first clicking the Select button, then drawing the desired polygon and clicking Run. The result is a display of the parcels located within the selected area. If the Show Related Items option is selected, a small form appears containing a list of territory names and toponyms in the specified area, along with the names of individual property holders. The Show Related Items feature can be used to identify, for instance, individuals or families who held property in proximity to some given location or feature. In the example in Figure 24 a selection was made of the area around the village of Petit Run. The results show that there were 13 individuals who are known to have held property near that town. This information may spark an investigation into potential interaction between these individuals and their families.
In addition to queries that can be performed from the map interface, the system supports map queries based on information from the main Directory form and from the People form. Examples of how to query from these forms are shown in Figures 24 and 25. These query options allow a researcher to quickly view the locations or properties related to specific transcriptions and to view the locations of properties held by individuals or family groups.

Figure 24: Spatial query options from the people form. In this example property holdings for one branch of the Regis family are mapped. a) Holdings where the location confidence is low. b) Holdings where the confidence is medium.

Figure 25: Spatial query options from the transcription directory form.
Mobile Access

To facilitate collaborative research, an important design consideration of the system is the ability to access and maintain information from mobile devices (Figure 26). This capability allows collaborators who might be in the study area to see the information previously collected for the region. Also, remote collaborators can flag specific places and attach related images or documents. Since all data updates occur in real-time, information collected in the field is immediately available to all researchers working on the study. Mobile access also provides a mechanism for researchers in the field to ground truth the data by visiting locations and comparing the extracted information with what is visible at the site.

Remote access capability is also extended to users with access to web browsers (Figure 27). This makes it possible for any user with access to the internet to view and, optionally, contribute information in a manner similar to what is available for mobile devices. In addition,
users with a more advanced background in map development can also develop custom maps to suit specific needs.

Figure 27: Browser interface which allows remote collaborators to view and contribute data.
Underpinning the user interfaces described above is a series of data services arranged in tiers (Figure 28). The foundation is the database tier which is made up of two databases, both of which are hosted using Microsoft SQL Server 2014. The first database is an ESRI format Geodatabase which stores geographical information and attachments that are used for map production. The second database is a custom database that stores transcription related details and ancillary data used to support the user interface (Figure 29).

Information from these databases is accessed directly by the two higher level service tier processes. The Geodatabase is accessed by an ESRI ArcGIS 10.3 service process making the individual features in the database accessible to client processes as ESRI Feature Services. There are nine feature service layers available for remote access through this mechanism. The Transcription Database is accessed using a custom developed service that is hosted under a Microsoft Information Services 8 (IIS) process. The custom service application programming interface (API) exposes 99 separate functions that support adding, editing, and querying.
transcription data. The custom API uses the Microsoft Window Communication Foundation (WCF) protocol. Both the ArcGIS service and the custom API run under IIS support web access. By implementing these services, the architecture allows users at remote locations to access the underlying data over the internet without requiring direct database access. This layer of abstraction provides greater flexibility and adds an additional layer of security to the system.
The user interface described in the previous sections of this chapter uses both the ArcGIS 10.3 services and the custom API. The ArcGIS services are used to support map displays, map queries, and the management of attachments. The custom API is used to support all other data access and manipulation functions such as searching for, and editing, transcriptions.

Mobile and web browser access to the system is routed from the ArcGIS 10.3 process through ESRI’s ArcGIS.com portal. The portal provides two functional benefits: the first is that it creates an additional layer of security which can be configured to only allow access for authorized users. The second benefit is that the portal allows the services published by this project to be found on the ESRI site using keyword searches thereby making them available to a much broader audience that would possible otherwise.

**Custom Coding**

The custom software for this project was developed within the Microsoft Visual Studio 2015 development environment using the C# programming language. The custom user interface is comprised of ten separate forms requiring 11,904 lines of programming code. The 99 functions making up the custom API required 2,647 lines of code.
CHAPTER 4 – The Case Study

Selection of the Study Area

The goal of this thesis is the development of an integrated and comprehensive system that allows researchers to extract historical information embedded in unstructured historical texts that can be analyzed and explored in a spatial context. The goal is to make information that was previously obscured due to its physical media and format accessible for scholarly research. While off-the-shelf GIS applications have been used to research historical events for which there is readily available structured data, this project requires an historical setting for demonstration purposes that is devoid of that resource. The challenge for this study is to extract, integrate, and expose information that otherwise would have remained elusive. Thus the selection of an appropriately challenging case study is required. A region and historical timeframe that satisfies this requirement is sixteenth-century St. Vincent; one of many small communities that were part of the background to events of larger consequence unfolding throughout major centers in Europe at the time. The available texts for St. Vincent are limited almost exclusively to formal recordings of property transactions or social contracts.

Along with historical documents, there are the physical artifacts produced by human activity, many of which can be seen in the relict landscape. There are the remains of stone walls that were constructed to demarcate farm plots or grazing areas, and canals that were constructed to bring water for use in agriculture and the home. Evidence from the Roman period can be found in the remains of residential structures, burials and a bridge crossing the Torrente di Cillan, or Cillan Creek, below the city itself. All of these landscape features contribute pieces to a story about population, status, and subsistence in St. Vincent; in essence the things that constitute the societies of the past. While there are no contemporary maps of the region, cadastral
maps of the area were produced in the late nineteenth century by royal officials acting on the authority of the Italian state. These maps contain detailed listings of the individual place names, parcels of land, and structures as they existed at that time. These maps are now available in digital format as scanned images. The maps and their accompanying documentation provide an important link between the physical features as they currently exist and the place names that remain from the sixteenth-century primary source materials. Because the texts and maps were created at different times and for different purposes, there is considerable uncertainty encountered when trying to create linkages between them. Several classes of uncertainty exist for these materials including inconsistencies in the original record, ambiguous person and place names, and lack of correspondence between the maps and the text sources. Adding further complication is the fact that the sixteenth-century records are written in French and Valdôtain (the regional patois), while the nineteenth-century cadastral maps were mainly in Italian. These factors in combination offer significant challenges to Historical GIS yet at the same time provide an ideal testbed for the current research application.

**St. Vincent**

Constructing a picture of St. Vincent as it existed in the sixteenth century is an exercise in assembling bits of information from sources separated in time by over 500 years. From these elements it is possible to tease out the place as it was known to its inhabitants. The earliest sources of direct information are notarial and administrative records concerned primarily with property transactions, legal disputes, and marriages. The unfortunate situation here is that there are no extant maps for the area from the sixteenth century or any descriptions that would give insight into the locations of the events that were recorded. What are available are place names, or toponyms, which are in many cases informal identifiers based on local topography or prior
ownership. The place names used in the sixteenth century, however, appear to display a remarkable durability over the years. Almost 400 years after the notarial recording of events, a set of maps was produced for the region that reflect many of the same toponyms appearing in the earlier records (Regione Autonoma della Valle d'Aosta, 2007). Importantly for modern research, these maps created between 1892 and 1897, are accompanied by listings of individual parcels of land and individual houses. One of the primary tasks then for the current research is to cross reference the toponyms from the earlier text sources and the maps in order to link events listed in the sixteenth century records to locations that were mapped centuries later.

Though we have hand-drawn maps of the region they are not geographically located. Through the use of GIS it is possible to combine the nineteenth-century maps with satellite and aerial imagery to produce georeferenced maps on which features, such as individual houses, can be located on the landscape. It is possible to establish the locations of stone walls that delineate field boundaries and trace the paths of canals, traces of which are still visible. But even with the use of modern imagery there are still limitations. For example, there may be features such as canals shown on the nineteenth-century maps but the present day aerial view of the same area is obscured by forest growth. Overcoming these limitations to produce a reliable representation of the built landscape of the sixteenth century is one of the primary challenges of this research, because it is a preliminary but critical step in the process of developing an exploratory Historical GIS that allows researchers to uncover spatial relationships between individuals, families, and the land.

Perhaps one attraction of sixteenth-century St. Vincent for an historian is the lack of previous research on the area (Vester, 2014). While this may seem somewhat counterintuitive, the situation offers new territory for investigation and presents an excellent test bed for the
application of a variety of research techniques. Fortunately, while there have been no formal articles or books published specifically on St. Vincent itself there is a great deal of documentation on the house of Savoy, under the recent heading of Sabaudian Studies, and on the Valle d’Aosta region as a whole (Ferrero, 1910; Vester, 2013b). This information can be used by extension to offer insight into the influences on St. Vincent and its political, economic, and social organization.

The sixteenth-century town of St. Vincent and surrounding communities are located in the Aosta Valley, or Valle d’Aosta. Archaeological evidence shows that the area was inhabited as early as the beginning of the first millennium BCE and the village was established as a Roman settlement by the first end of century BCE (Daudry, 2015). The Valle d’Aosta is now an autonomous region in the Alps of northwestern Italy, though in the sixteenth century it was a largely self-governing duchy within the composite state of the Duchy of Savoy (Vester, 2013b). This region had been part of the lands controlled by the house of Savoy for several centuries and was given ducal status by Emperor Sigismund in 1416. It must be understood, however, that ownership of a territory in the form of a duchy was quite different from what is understood today in terms of nationhood. During the sixteenth century loyalty and affiliation were based on relationships between individuals rather than to governments. In fact, the concept of government as it is understood today did not exist in the region during this period (Vester, 2013a) (Raviola, 2013). Instead, government was performed within a relatively informal hierarchy of elites involving administrators at the local village level and going up through representatives known as the Conseil de Commis (Vester, 2000). The underlying criterion for leadership was family affiliation and its attendant access to wealth, primarily in the form of land holdings. So, while
being under the domain of the Savoy dynasty implies formal control, the local elite maintained a
great deal of autonomy in its day-to-day administration.

Because access to power was through family affiliation, it was important for families to acquire wealth in the form of land holdings and also rights to resources such as water for irrigation. Given that wealth, and therefore power, were associated with family, the ability to establish strong alliances between families was an important consideration. Because of the importance of land ownership and family alliances keeping detailed records of transactions became standard practice and local notaries were active in recording sales, marriages, inheritances, and other assignments of wealth and power. For the modern researcher this has resulted in a rich archive of notary records that offer detailed insight into the names and places involved.

**Texts**

The notarial records for St. Vincent, which include listings of property transfers, marriages, deaths and other transactions, are currently housed in the *Archivio Notarile Distrettuale di Aosta* in the city of St. Vincent (Vester, 2014). Written in the local sixteenth century patois, these documents have gone largely unexamined over the centuries. The difficulty in translating these documents is not simply that the language has changed over time but also that the form of handwriting, and more specifically the shorthand of the notaries, has evolved. Over the past twelve years there has been ongoing work to analyze these documents resulting in a transcription of transaction details for portions of the region and, in particular, for the Regis family. So far one of the approximately 30 registers has been completely transcribed along with portions of several others. This transcription is one of the key components that allows us to
establish family relationships and to associate specific geographic locations to events from the period.

The translation process itself requires careful thought. A casual translation, as one might encounter in a foreign language study guide, is not sufficient for historical research. For example the French phrase “un champ de 1 hectare en taille” could be translated as “a field, 1 hectare in size” or as “a meadow, 1 hectare in size” or simply as “1 hectare of land”. While all of these translations indicate that a specific quantity of land was exchanged and that there was an implicit or explicit associated value, there are geographical and social elements that may be lost. In a twenty-first century setting, parcels of land are often purchased without the need to record previous function. In an agrarian society, such as sixteenth-century St. Vincent, the type of land is of great significance and as a result the translation of the word “champ” in the example above takes on additional significance in our understanding of the participant in the transaction. “Pre” in the context of St. Vincent of the period indicates a meadow that was used to grow grain for livestock. If, on the other hand, the word “champ” had been used instead, it would imply that the parcel was a “field” that was cultivated and used to grow wheat or oats. The implication of the first is that the purchaser may have owned sufficient livestock to require additional feed crops. The purchase of a “champ” on the other hand implies that the purchaser had a need for additional land for growing wheat that could be turned into bread for the family or for sale. In a situation such as this a researcher would look to other details in the transaction that might imply how the grains might be put to use.
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 the tax surveys could be ambiguous, the location of properties on a map left little room for hedging and helped to ensure that every piece of property was accounted for. Unfortunately no contemporary sixteenth-century maps are available for St. Vincent or its immediate surroundings.

The earliest maps that we have for St. Vincent were created between the years 1893 and 1897 (Regione Autonoma della Valle 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 well developed cartographic skills, with individual parcels and structures delineated and numbered. Each of the detailed maps was accompanied by an index of the territory (frazione) and toponym (toponimo) in which each parcel was located. These documents, or foglio, indicate that the surveys themselves began ca 1892 and continued at least 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.

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. The link from the maps to the current landscape is based on the ability to georeference the maps to satellite and aerial imagery using GIS.

Figure 30 shows an example of a source map. The label on the top left edge indicates that this particular 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 by staff at the archive from scans of the original hand-drawn maps.

Structures are identifiable because of their size and geometric shapes and also because of the shading which is applied to their boundaries. The only indication of function for any of the
structures is a cross that appears within the boundaries of some, implying perhaps a religious function. 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 interest of the survey itself was beyond that of the local community. However, it would seem that even for the purposes of taxation an indication of function may have been expected.

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. At that point one train station and four toll booths were deemed significant enough to be labeled.

A close-up of a section of this same map (Figure 31) reveals some of the detail that is available and some of the visual indicators that were used to determine feature types. Land parcels are indicated by thin borders while structures are identified by thicker lines with internally shaded borders. Among the structures, there is one in the right central section of the

Figure 31: Detail of map 6. Land parcels are indicated by thin borders, 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.
map that has a cross drawn in it. This structure is interpreted to be a church or some other religious building although in this case it is left unnamed. All of the parcels and many of the structures are numbered with values that correspond to the associated map index. The indices, in turn, indicate the territory and toponyms for the various ranges of numbered items. In this example, it is possible to clearly identify the 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; the Strada di Sape and Strada da Perriere.

This particular map is unique among the set of 67 in that it is signed; there is one signature of a Giovanni Croveri, one of Supr Luigi Coreggiari, and two signatures for Giovanni Reales. It is not possible to determine from the available scanned images if these signatures are contemporary with the maps themselves. We also have no record of who these individuals were; what their role was in a survey team, or whether they local dignitaries or people working in the archives who were simply practicing their flourishes. This is the type of small mystery that comes up on occasion that adds a touch of unexpected intimacy and intrigue to historical research.

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 32). A subsequent comparison of territory and toponym names seems to confirm that many of the names on the maps do in fact accurately locate
the 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).

For this research, the real key to these maps is that the numbered parcels correspond to 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. 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 working with the sixteenth-century notarial records. The most obvious issue is that the ledgers for the maps one through eight and also map ten are missing altogether. 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 difficult to decipher names, reducing the reliability of matching place names from the sixteenth-century texts with the names from the surveys. The next step is to examine how the system that has been created can be used to uncover the information hidden in the historical materials from sixteenth-century St. Vincent.
CHAPTER 5 – The System in Practice

Application of HGIS to sixteenth-century St. Vincent

Off-the-shelf GIS can be used to produce maps and display spatially related attribute data and attachments. It can also be used to identify features within user selected areas or features that match user specified criteria. As noted above, however, GIS functionality is predicated on the availability of well-defined and well-structured data. The goal of this thesis was to develop a unique spatial information system capable of handling unstructured historical and geographical data. The system builds on traditional GIS mapping and querying capability through the development of an integrated information management system that facilitates the use of unstructured data for the production of spatially aware knowledge. Rather than producing fixed visual representations that present the user with a pre-defined maps, this system involved the user directly in the information gathering and knowledge production process. The only “maps” that are produced are those that answer questions of interest to the researcher. The real value of this system, however, lies in what it allow a user to discover about the historical geography of a place and in this case St. Vincent.

In this case study, the system with its data analysis and user interaction capabilities, are used to contribute to a better understanding of familial relationships and property holding strategies in sixteenth-century St. Vincent. This understanding and insight is based on information taken from the transactions recorded in the source materials as well as from information gleaned from the wording of the records and the manner in which the participants are referenced. Using this system, place name information from the texts can be analyzed in conjunction with spatial information from the reconstructed landscape to derive the locational
significance of landholdings based on features such as proximity to resources, transportation routes, and population centers. Familial and social relationships can be analyzed based on the roles of the participants identified in the source materials. In the study of social and familial in a small rural community the problem of ecological fallacy can lead to a notion of a common sense of purpose and predetermined patterns of interaction (Butlin, 1993, p. 69). This research provides details and new insights that can help the researcher obtain a more informed viewpoint.

**Historical Materials**

So, how specifically can the functionality described in chapter three be applied to the case of sixteenth-century St. Vincent? The following sections look at how the system can be used to address some of the important topics of historical interest.

**Space and Place**

Location information in the notarial records is limited in general to the use of toponyms. The transcribed sixteenth-century notarial records of St. Vincent do not suggest the development of a standardized place naming convention, particularly one based on numbered addresses. Thus transactions between individuals that crossed social and cultural boundaries and the recording of property transactions required a system that could be understood unequivocally regardless of world-view. In St. Vincent the parties to the majority of the transactions seem to have been individuals with similar geographical, if not social, background. These individuals would have grown up in the region and would have developed the same innate understanding of places and the geography of the region. Because of this, an informal naming structure was sufficient.
In St. Vincent the naming conventions in use seem to be reduced and simplified from those observed by Smail (1999) in Marseille. The St. Vincent records are dominated by place references based on property function, proximity to a village, prior ownership by a recognizable individual, or topographical feature. Only occasionally are there references to a ‘street’ and then only as descriptions such as “rue dessous” meaning “road below” or “rue dessus” meaning “road above”. An example of this can be seen in a transaction recorded on February 14, 1566 between Vincent Granges and Thonine Dartha daughter of Guillaume de Lettry and the reference to a property as the house of Bornel (“mayson du Bornel”) located in the town of St. Vincent on the “rue dessus” or the upper road. There is also no mention of administrative boundaries other than occasional references to properties located in “the bourg” or city. This term is generally used to reference St. Vincent itself but occasionally is used to reference smaller villages as well.

Given the somewhat tenuous but critical link between the place names, or lieudit, of the sixteenth-century patois and the toponimo of the map ledgers, it is clear that a mechanism was needed to facilitate navigating the ambiguity and fuzzy name matching based on similar spellings rather than exact matches. Given the relatively small number of source texts and the fact that the sixteenth-century patois was being compared to Italian, the development of an automated process for disambiguating place names such as that used by Giusti (Giusti et al., 2007) was not feasible. A much greater linguistic investigation into sixteenth-century patois, known as Valdôtain, would be required before useful results could be produced using this method. The approach taken in this study is based on the development of a compiled thesaurus of place names. However, rather than applying the thesaurus to link works with similar semantic meaning, it is instead used to link spelling variants to referenced similar places. In practice the thesaurus is designed to be populated as part of the exploratory process of examining the source
documents. As new place names are encountered in the texts, preliminary searches can be performed against the nineteenth-century map toponyms based on phonetic spelling. These coarse search results can be refined based on a spatial analysis and on possible clues found in the texts. Once synonyms are identified they can be added to the thesaurus and used in subsequent queries. For example, in a transaction in register 131 dated July 7, 1567 there is a reference to a toponym spelled “Le Grenya”. An exact search of the map toponyms based on this spelling returns no matches. However, experimenting with letter substitutions and various letter pattern matches produces three toponymic matches located in four distinct territories with the spellings “Gragna”, “Grania”, and “Gragnas”. With these toponyms the researcher can perform a spatial query to identify possible locations. The results indicate that all of these toponyms are located in the same local area. This gives the researcher some confidence that the names reference the same geographic location. With this knowledge the thesaurus can be updated with entries that link “Grenya”, “Grania”, and “Gragnas”. The dynamic character of the thesaurus allows the system “knowledge” of the region to continually grow as new transcriptions are added.

In some cases, however, it is possible to have matching thesaurus entries that do not necessarily refer to the same location. Take for example the toponym ‘Molin’ from another entry in register 131 from the same date. Performing a search of map ledger entries turns up 29 possible matches with names like ‘Moulin, ‘Le Moulin, and ‘Molinet’, and ‘Campo du Moulin’. In this case the name is actually more of a descriptive term rather than a formal name. Moulin means mill, so any location with a mill could potentially have “Molin” as part of its name. This situation is similar to a generic modern place name such as Sandy Beach which could refer to one of many places that are named based on the characteristics of the local topography. However, adding all of the variants into the thesaurus is still very useful for narrowing down
Because of the integrated nature of the system, when the name “Molin” is encountered in the texts, the ambiguous toponyms can immediately be plotted and their spatial relationship to other known properties can be used in many cases to identify with some degree of confidence which one is appropriate. The selection of a toponym for a particular transcription may be based, for example, on proximity to other known properties owned by a buyer or a seller. In summary, the systems capability to match names, in combination with its integrated mapping functionality, provide a unique spatially enabled exploratory tool set that can be used to develop a deeper understanding of the physical and social landscape of St. Vincent.

**People**

As transcriptions are entered into the system an ever growing set of family lineages continues to develop, providing information on both consanguineous and affinal relationships. Within the currently transcribed notarial records, there are 145 distinct surnames although these may translate into a larger number of individual family groups. There are also references to 80 individuals who were referred to as being deceased at the time of recording. Death is indicated by the use of the word ‘feu’, for example “Henry son of feu François Ambrosin” in an entry dated July 11, 1567. The use of the word ‘feu’ appears to have been a broadly used convention, however, there are cases where an individual may be referred to as “Henry son of feu François Ambrosin” in one transaction and subsequently as “Henry de François Ambrosin” in another. It is unclear from the transcriptions if this was due to style preference on the part of the particular notary or if there was some time period or family event that triggered the qualifier to be dropped.

Naming for both men and women followed a generally consistent pattern that was based on paternal lineage. Individuals were generally identified by their given name along with that of
their father and in most cases at least one additional level of paternal ancestry. For example, the name “Vincent de Jean de Mathieu Chadel” represents an individual named “Vincent Chadel” whose father was named “Jean” and whose grandfather was named “Mathieu”. It is rare to not have an ancestor listed although this is almost exclusively the case when an individual is identified as being the notary, as in a transcription dated February 4, 1566 which references a “Philibert Loczon“. The system allows these implicit indications of dates of death to be recorded for individuals and used for subsequent analysis.

Status and titles also appear occasionally in the records; the most often being the title “Noble”. Tracing the appearance of “Noble” may provide some insight into when an individual made the transition from relatively common status into the higher echelons of society. For example, the following references to Vincent Regis seem to indicate that at some point between November of 1582 and April of 1595 he adopted such a title:

- April 21, 1595: Noble Vincent Regis
- November 17, 1582: Vincent d’Humbert Regis
- June 3, 1579: Vincent d’Humbert Regis

In addition, titles of nobility were not exclusive to men as can be seen in an entry from March 23, 1587 which refers to “Noble Marguerite widow of Guillaume Regis”.

A complicating factor when trying to identify specific individuals is that it is very common in the registers that spelling variations occur in first names. Many of these variations are simple substitutions of phonetically similar letters such as “y” for “i” and some involve the additional of silent characters such as “h” or a space. It is not clear if these variations were due to the predilection of a particular notary or if there were other factors involved. Some of the common examples are:
<table>
<thead>
<tr>
<th>Name</th>
<th>Spelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jehan</td>
<td>Jean</td>
</tr>
<tr>
<td>Jehanantoine</td>
<td>Jean Antonie</td>
</tr>
<tr>
<td>Antoine</td>
<td>Anthonie</td>
</tr>
<tr>
<td>Bartholome</td>
<td>Bartolome</td>
</tr>
<tr>
<td>Francois</td>
<td>Francoys</td>
</tr>
<tr>
<td>Nicolas</td>
<td>Nycolas</td>
</tr>
</tbody>
</table>

This system is designed to allow the searching and matching of these ambiguous spellings using the same ‘Thesaurus’ approach as is used for place name matching. And, similar to place names, name spelling variations can be dynamically updated as they are encountered and then used in subsequent queries.

There are also, on occasion, descriptive references to individuals in the records as in:

“Marte, wife of Boniface d’Anthoine du Crest, granddaughter of Pierre Bertrand Mistral, daughter of Perronette (Mistral’s daughter), and of Johannet Richerand”

This type of extended description, while somewhat convoluted, is very useful for reconstructing family lineages in a single sentence. In this example we are given a three generation lineage beginning with Pierre Mistral through his daughter Perronette then to his granddaughter Marte. We also know that Marte’s father was Johannet Richerand, although we assume with some confidence that Perronette and Johannet were in fact married. Finally, we know that Marte was married to Boniface who was the son of Anthoine du Crest.

The system allows all of these pieces of information gleaned from the notarial records to be recorded, resulting in a dynamic and growing set of family lineages that are linked to transactions and can therefore be spatially and temporally located. This information can be used to examine family land-holding and marriage strategies.
Inheritance

Risk avoidance was an important consideration in the lives of peasants and small landholders in the Alps during the sixteenth century. A key component for maintaining stability for the individual, and more importantly the family, was the development of a careful strategy of accumulating and holding land and in the most beneficial locations (Siddle, 1986). Maintaining stability often meant maintaining a number of small plots in relative proximity, shared among agnatic groups. This strategy allowed family members to share resources and labor toward a common goal. With the death of a landholding member of a family (generally a male), land was distributed to his heirs in a manner that would maintain the stability of family holdings for as long as possible. Cases of inheritance distribution in the notarial records indicate that property was generally distributed equally among (primarily male) heirs and that individual parcels were not distributed wholly to individual sons but rather individual pieces of property were held in shared ownership. For example, in a transaction dated July 11, 1567 the three sons of the deceased Francois Ambrosin were each given one third ownership of a house and one third ownership a parcel of land. This example indicates that the pattern of property inheritance was one whereby ownership was distributed to all surviving male heirs to be shared in common.

Based on the transactions in register 131, it appears that families in St. Vincent practiced a pattern of shared inheritance although this single source of evidence does not exclude the possibility that patterns of partible inheritance or primogeniture existed as well.

The integration of property holdings with family lineage information allows these types of strategies to be evaluated. In addition, an attribute of each transaction stored within the system is its type. One type of transaction was a division upon the death of a property holder. Using this
piece of information, property transactions can be queried by individual, family group or affinal relationship thereby allowing system users to research the property strategies being used.

**Land Use**

The capability of the system to integrate the notarial record data and the maps offers insight into the society and social interactions of the region at the time. The system allows property information to be located in a spatial context making it possible to see the influence of terrain and elevation on the lives of the inhabitants. It is possible to see where different numbers of transactions occurred and thereby imply land values. Along with offering information about the people and society, this combined information may be used to extrapolate information about climate and land use patterns. Transactions involving vineyards, for example, were located on land that is primarily south facing with minimum obstruction of the morning sun and at elevations that ranged from between 400 meters at the valley floor to as high as 1,000 meters. A review of the aerial imagery of the region as it is now indicates that this spatial pattern appears to be similar to current conditions.

Transactions involving land at the highest elevations refer primarily to grazing lands based on the use of the term “champ” or field in describing the property. At lower elevations there are higher occurrences of the term “pre” or “meadow” implying larger reliance, or potential, for raising crops. It is estimated that the annual yields for grasses decreases by 40% per 1,000 meters of elevation gain (Mathieu, 2009, p. 223). With the number of transactions currently in the system it is not possible to determine if this is reflected more widely in the records of St. Vincent, however, a preliminary visual review of the distribution of property types indicates that some differentiation can be discerned based on elevation.
An interesting example of how the system facilitates the interpretation of the notarial records can be seen in a series of transactions between members of the Ambrosin family on July 11, 1567. There are a total of 47 transactions each identifying individual pieces of property. Seventeen of these transactions specifically reference property using the term “pre”, implying some degree of wheat production. There is also a transaction on that date in which a mill is exchanged. In combination, these transactions indicate that the Ambrosin family was involved in both the growing and milling of grain, although it is not possible to get an indication of how much of their milling operation was for their own grain production rather than that of others. Further exploration of the Ambrosin family indicates that along with the “pre” and mill transactions, other entries include fields, vineyards, trees and various buildings. In combination, these records can be used as a gauge of the wealth of the family at the time and as an indication of the way in which they diversified their holdings to ensure the stability of the extended family.

In addition to identifying where various types of property existed, the system also helps highlight what may be missing from the notarial records. During the sixteenth century there was an active timber industry in the Alps and with timber slides constructed to aid in timber transport (Mathieu, 2009, p. 122). Transactions in which trees are involved are relatively common in the transcribed documents, though there is little indication that they were used for trade for many of the entries refer specifically to chestnut, apple or other fruit bearing types. As in other cases, additional transcriptions will be required to learn more on this topic, however, the integration of the database with the exploratory interface allows this type of question to be examined as more records are transcribed and entered.
Use of Additional Sources

It is important to reiterate that the system does not rely on a specific type of source document. Other types of records, such as birth records, death records and tax survey data can also be ingested. The notarial records alone offer point-in-time indications of property transfers. Using property transfers, it is possible to say for instance that a given individual came into possession of a given piece of property on a given date and that another individual relinquished possession of that property. Without a baseline to work from, it is not possible to reconstruct the overall holdings of these individuals, much less develop a picture of the extent of their family holdings. Fortunately, even though the region maintained a semi-autonomous status, the influence of the Duchy of Savoy did not go unnoticed in the Valle d’Aosta during the period. The result of this was the creation of property ownership registers known as *reconnaissance* records, or accountings, of holdings for tax purposes. These surveys, in theory, contain listings of all property holdings for the region at the time and can therefore be used as a baseline from which to reconstruct an individual’s or family’s holdings at any point in time. This reconstruction can be accomplished by applying purchases and sales, either prior or after the survey, to these records. The results, however, must be used with caution because for the small landholder these registers defined their tax burden (Federzoni, 2013; Vester, 2000). The desire to reduce this burden led to a situation where the small landholders would attempt to understate their holdings and thereby reduce their tax payment (Brucker, 1971, p. 23).

The dates of the transactions in the notarial records can also be used to bracket the birth and death dates of individuals. If a person acquired a piece of property on a given date it can be assumed that the person was born prior to that date and died after that date. If a person was listed as ‘feu’ it can be assumed that he or she was deceased at the time. As new transcriptions are
entered into the system this information can be recorded. Unfortunately this leaves a lot of
uncertainty as to actual birth and death dates. More explicit information is available, however, in
the form of birth and death records from the period. The system supports the inclusion of these
records though so far none of these have been transcribed. Additional information regarding the
age of individuals at the times of specific transactions may be useful in developing an
understanding of topics such as the age at which property could be acquired and when, prior to
death, it may have been distributed to children.

Additional Reporting Capability

As the title of this thesis implies, this is an exercise in the use of unstructured material
within the highly quantitative and structured environment of a GIS. A key feature of the
architecture of this system is that it is based upon the use of a comprehensive data model that is
implemented within a SQL based database. This design feature has the benefit of providing an
ability to produce summarized statistics that can expose patterns of actions and events that may
otherwise go unnoticed. These patterns can be related to the information recorded explicitly in
the records, such as the number of sales of water rights, or implicitly, such as the number and
types of references to women. The possible queries can range from reporting the total number of
transcriptions in the system to a breakdown of the role of individual parties to transactions based
on gender or the frequency of occurrence of various types of property (Appendix A).

A variety of off-the-shelf software applications can be used to develop these types of
queries including widely used ones like Microsoft Access and Excel. In addition these, the
database used to implement this system has a built-in browser based component that is designed
to support ad-hoc end-user queries and reporting. These query features allows researchers to
extract information based on their goals and based on questions that arise during their use of the system.

In aggregate, the integrated set of data analysis and user interaction capabilities developed for this thesis can be used to contribute to a better understanding of familial relationships and property holding strategies in sixteenth-century St. Vincent in ways not possible using a standard GIS. By taking advantage of the ever growing knowledge base that is augmented with each new transcription, a picture can be formed of sixteenth-century St. Vincent with ever increasing detail.
CHAPTER 6 – Future Research Direction and Conclusion

Future Directions

Even though it may be possible to claim technical success with the work that was performed for this paper there are many areas for improvement and a great deal of follow-on historical research still ahead before a final verdict can be given. There is no technical reason that this system cannot be applied to times and places other than sixteenth-century St. Vincent, yet it is not possible to determine its effectiveness or applicability at this time. Additional case studies using a variety of types of source materials will need to be tested before a conclusion can be reached. Also, at the time of writing the systems has only been used in a controlled setting by a very limited number of users. Because of this it has not been fully exercised and its shortcomings can therefore not be adequately evaluated.

Based on initial testing and on the anticipation that the system may one day be used for actual production, several functional and technical enhancements have been identified. These include:

Functional Enhancements

- Enhance the Patois word list
  - Improved automated translation capability that will help expand the user base

- Build a comprehensive thesaurus
  - Improved automation of person name and place name identification

- Build a rule based name matching process similar to Giusti (Giusti et al., 2007)
  - Improved automation of person name and place name identification

- Create additional linking capability between transcriptions, people, and maps
  - Enhanced exploratory capability that may lead to additional insights
• Build support for evaluating data base on gender.
  o Enhanced understanding of gender roles

• Develop enhanced symbology and map presentations of toponyms and ownership
  o Enhanced exploratory capability that may lead to additional insights

• Address the time component
  o Enhanced ability to examine the sequence of events leading to a better understanding of cause an effect.

• Develop spatial analytics capabilities
  o Examine the implications of factors proximity to resources, amount of property owned, property per family member, geographical distribution of various property types

• Develop a capability to explore intra- and extra-family interaction
  o Pattern identification – spatial and social

**Enhanced Support for Historical Geography and Historical Research**

• Include tax surveys
  o Development of a better understanding of total property ownership

• Include birth records and death records
  o Development of more precise family structures
  o Develop estimates of regional populations

• Examine distribution of inheritance between siblings
  o May provide baseline for establishing total family holdings

• Examine holdings and transaction patterns among families
  o May contribute to a better understanding social stratification

• Continue to incorporate additional records as they are transcribed
  o Additional information to fill in knowledge gaps and clarify details

• Evaluate the applicability of the system to other historical settings
  o Contribute to historical research in other settings.
Technology Enhancements

- Plan for scaled-up multi-user use
  - Will allow for more rapid transcribing
  - Collaborative process may generate new geographical and historical insights

- Build in authentication and use SSL connection
  - Provides greater information security and stability
  - Reduces the possibility for data corruption

- Define and set up ‘crowd sourcing site’
  - Will allow additional ancillary data to be collected for analysis

- Build a editing application for mobile devices
  - Will support greater opportunities for system use including on-site data collection and exploration

- Monitor system/user interaction
  - Will provide information that can help identify functional, subject matter, and technological enhancements

- Develop standards based data interchange capability
  - Will open up system access to broader community of researchers

Conclusion

This paper has discussed the development of Historical GIS as a technology and as a tool that can be applied to contribute to a deeper understanding of the spatial context in which historical events took place. The research has also addressed the issues surrounding the use of unstructured historical data and maps. Where the use of GIS is normally contingent upon the availability of structured, tabular data that is amenable to use in computer systems, this research evaluated methods for ingesting unstructured texts and identifying embedded information. It has also demonstrated the use of hand-drawn maps where GIS normally relies on the availability of digital maps built around standardized coordinate systems. With the challenges of using
historical material address, this research developed a mechanism for supporting knowledge extraction. The resulting system was built around a knowledge repository that is tightly integrated with a spatial query and visualization interface. The sixteenth-century town of St. Vincent was used as a case study to evaluate the system’s ability to enhance the exploration of historical family lineages, events and places. However, because the use of the system is still in its very early stages, a final verdict on its success in enhancing the knowledge of St. Vincent is still pending. Even so, early testing shows encouraging results, and the system has sparked interest among those who have seen it in operation.

This paper has demonstrated that unstructured materials can be successfully accessed and analyzed as an information source within a Historical GIS. However, this solution sits beyond the more generic use of off-the-shelf GIS as predominantly used in Historical GIS. The research has shown that historical insights can be gained by developing an exploratory spatial system in which the user is an integral part of the process; acting as simultaneously a researcher and contributor.
REFERENCES


Ferrero, F. G. (1910). The Valley of Aosta: A Descriptive and Historical Sketch of an Alpine Valley Noteworthy in Story and in Monument.

Fu et al. (2014). Automatic record linkage of individuals and households in historical census data. International Journal of Humanities and Arts Computing, 8(2), 201-225.


APPENDIX A

Example Query Results

<table>
<thead>
<tr>
<th>Year</th>
<th>Transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1561</td>
<td>1</td>
</tr>
<tr>
<td>1562</td>
<td>2</td>
</tr>
<tr>
<td>1563</td>
<td>3</td>
</tr>
<tr>
<td>1564</td>
<td>120</td>
</tr>
<tr>
<td>1565</td>
<td>83</td>
</tr>
<tr>
<td>1566</td>
<td>157</td>
</tr>
<tr>
<td>1567</td>
<td>85</td>
</tr>
<tr>
<td>1568</td>
<td>34</td>
</tr>
<tr>
<td>1569</td>
<td>95</td>
</tr>
<tr>
<td>1574</td>
<td>2</td>
</tr>
</tbody>
</table>

Transactions in Register 131 by year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acquirer</td>
<td>Relinquisher</td>
</tr>
<tr>
<td>1549</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>1561</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>1562</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>1563</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1564</td>
<td>77</td>
<td>12</td>
</tr>
<tr>
<td>1565</td>
<td>38</td>
<td>70</td>
</tr>
<tr>
<td>1566</td>
<td>52</td>
<td>28</td>
</tr>
<tr>
<td>1567</td>
<td>38</td>
<td>57</td>
</tr>
<tr>
<td>1568</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>1569</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>1574</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Individuals mentioned as parties to transactions by gender.

<table>
<thead>
<tr>
<th>Property Type</th>
<th>Transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meadow (Pre)</td>
<td>83</td>
</tr>
<tr>
<td>Field (Champ)</td>
<td>78</td>
</tr>
<tr>
<td>Cheyney/Chenevir</td>
<td>35</td>
</tr>
<tr>
<td>Trees</td>
<td>35</td>
</tr>
<tr>
<td>House</td>
<td>34</td>
</tr>
<tr>
<td>Arbor (Thopiage)</td>
<td>27</td>
</tr>
<tr>
<td>Vineyard</td>
<td>25</td>
</tr>
<tr>
<td>Tree - Chestnut</td>
<td>21</td>
</tr>
<tr>
<td>Water</td>
<td>20</td>
</tr>
</tbody>
</table>

Nine most frequently referenced property types.