Exploring the History of Venice: Relics, Records, and Relations

AN INTERACTIVE QUALIFYING PROJECT REPORT
WORCESTER POLYTECHNIC INSTITUTE

Submitted to:
Project Advisor: Fabio Carrera, WPI Professor
Project Co-Advisor: James Cocola, WPI Professor

Submitted by:

_____
Corrie Avila

_____
Elizabeth De Zulueta

_____
Sofia Mukhanov
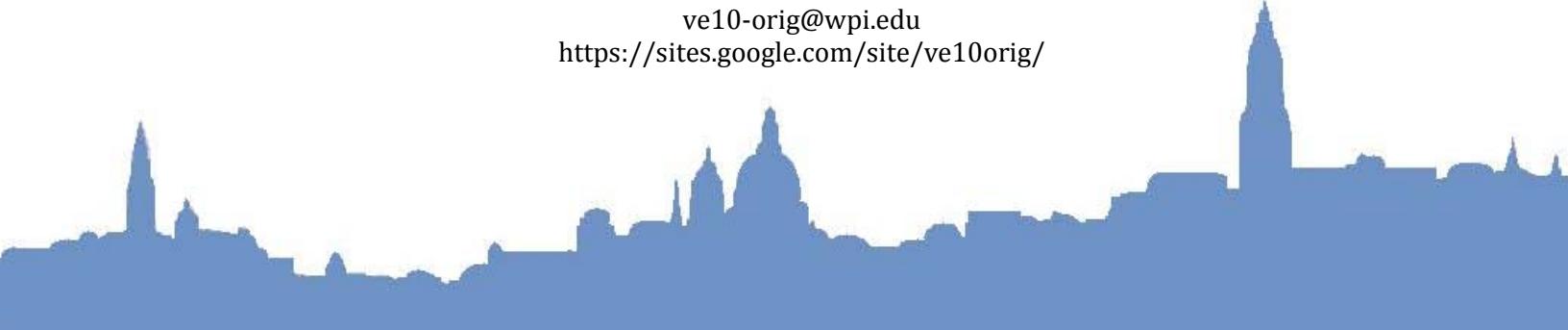
_____
Maarja-Liisa Sokk

Date: December 18, 2010

ve10-orig@wpi.edu
https://sites.google.com/site/ve10orig/

Authorship

Each member of our team contributed equally toward the completion of this project. Without the full participation and contribution of each individual, the following project would not have been possible.

Acknowledgements

Abstract

This project is comprised of three unique and independent components. The common thread which ties them all together is their inter-relationships with the city of Venice and understanding its historical roots. The first component promoted a collaborative transcription assistant tool, *uScript*, to facilitate research in the Venice State Archive. The second component enhanced archaeological research through the use of mobile applications to improve excavation data collection and promote previously excavated sites. The third component utilized historical data of the ancient Venets and created a plan to determine their origins using genetic genealogy. Each of these components provides validation of the others and fills in gaps to form a complete picture of the ancient city of Venice.

Table of Contents

List of Figures

Executive Summary

Venice is extravagant and intriguing, as it intertwines the magnificence of the ancient city with the hustle of the new city. Much of the history of Venice can be found right within the city. The Venice State Archives hold miles of records and ancient documents that contain information from every aspect of ancient culture; from the actions of the Doge to the records of the common-people, and everything in between. Researchers are constantly delving into the vast information these documents contain, dating as far back as the year 1000. Similarly, archaeologists are digging into the city itself uncovering incredible amounts of fascinating artifacts and relics which give insight to the ways of the past. Additionally, recent efforts are being made using the analysis of DNA and genetics to uncover more information on the history of Venice and the ancient Veneti, whom we will be referring to as the Venets from now on to avoid misunderstandings. This genetic analysis is combined with evidence of the Venets' presence throughout Europe. This evidence includes traditions and customs including mythology and symbolism, as well as archaeological evidence such as urns, brooches, and figurines.

This project aimed to gain a better understanding of the history of Venice and the origins of the ancient Venets, by way of:

- Facilitating the development of the uScript transcription tool for archival research,
- Designing and demonstrating the usefulness of a mobile application for the collection of archaeological data, and
- Organizing and planning genetic and historical research on the ancient Venets.

History of Venice

Tools for Historical Research Through Ancient Documents

The Venice State Archive contains a plethora of ancient records and documents which portray the history of the city and its international relations. The expansive collection is constantly being used by those intrigued with the information it contains. To access the information within the archive, a researcher must request a specific manuscript and reference it within the Venice State Archive Study Room. The transcriptions and notes that they produce from this manuscript are purely their own; there are no copies left behind to reference and future researchers will not be aware of any research that has already been completed. Additional researchers wishing to access information on the same topic would likely request the same manuscript and produce a similar transcription, which he would then also file in his personal records. This process is extremely redundant and leads to a significant amount of usage of the already fragile manuscripts.

The students of Worcester Polytechnic Institute have created a solution to this foreboding issue. Previous years' project teams designed the program *uScript*, which is a transcription assistant tool to be used within the archives. This program allows for collaboration and sharing of transcriptions by storing them in a publicly available database. Along with entered transcriptions, this database also contains digital images of manuscript pages so that researchers can avoid touching the manuscripts directly, thus preventing further deterioration.

Our project team continued the *uScript* campaign by updating the uscript.org website, creating opportunities to advance the program code, and searching for funding in support of the program. The *uScript* website was restored, updated, and edited to better convey to objectives of the program. Internship and school project opportunities were created to facilitate the completion of the program's code. Additionally, funding request letters were written and sent to potential sponsoring agencies and individuals that may have an interest in the future of the program.

Tools for Facilitating Archaeological Processes

The current archaeological process in Venice is long and cumbersome. At each site an archaeologist must take numerous pictures, describe the scene thoroughly and fill out a series of forms. This procedure uses up valuable time especially at mundane sites such as work on septic tanks. Once the archaeologists complete all the paperwork they must compile all the information collected into reports, some of which become available to the public. Our project team has created two mobile applications, the first to ease the work of archaeologist in the field, and the second to promote archaeology in Venice to everyday people.

We designed the first application to facilitate the process of data recording during archaeological excavations. With the help of Venetian archaeologists, we researched the excavation process and the information that must be denoted on site. The implementation and use of our mobile application will simplify this process and allow for more information to be recorded. All the information entered into the application, such as pictures, descriptions, relations, and artifacts, will be stored in an online database that can be easily accessed from the office. Once coded this application will help minimize time wasted in the field for routine excavation sites that archaeologist are mandated to be present at.

The second application deals with the archaeological sites whose reports have been made public. This application uses the Layar platform to display basic information about each site on the world around the user through their mobile phone's camera view. For each of the six sites currently on the application users can view a picture of what the dig site actually looked like. The Layar platform also allows users to get directions to the sites directly off their mobile phone if they are interested in seeing what the site looks like after the excavation. This application is more intended towards individuals interested in learning about archaeology in Venice. As more reports

are made public, this application can be updated to include a much more in-depth account of Venice's history.

Tools for Genetic Research to Determine Origins of the Ancient Venets

Despite the amount of history contained within Venice, much of its ancestral past remains a mystery. Considerable archaeological and historical research has been compiled regarding the Venetians' predecessors, forming a number of theories about their origin. This study focuses on the influence of the Venets, throughout Europe. The regions we determined for genetic analysis are Brittany, Northern Wales, Germany, Czech Republic, Poland, Northern Turkey, and Slovakia. Although genetic analysis has been performed previously on a number of the current inhabitants of Venice, the results were inconclusive due to the mixing of DNA that occurs in large cities.

Our contribution to this research was establishing a plan for the collection of DNA within the aforementioned regions. We have formed agreements with collaborators at these locations who will collect 90 DNA samples from the regions of Brittany and Northern Wales, Northern Turkey, Slovenia, and the Veneto itself. These samples will be collected from the adult male population having at least 3 generations of family from the given testing region using Watman Sterile Omni Swab Kits. Once collected, these samples will be mailed to the Genographic Project Laboratories in Barcelona, Spain for analysis. The results will be compared to samples already in genetic databases including samples from Lusatia, or modern day Germany, Poland, and the Czech Republic, to hopefully find a genetic link between the peoples of these locations. These outcomes will be compiled along with historical and archaeological evidences to establish a better understanding of the Venets' origin.

1. Introduction

From international relations, to government policies, to personal ancestry, history is a guide in our everyday lives. Cicero once stated: "To be ignorant of what happened before you were born is to remain perpetually a child. For what is the worth of a human life unless it is woven into the life of our ancestors by the records of history?" Each of the 6.9 billion people populating this Earth is related. Spanned across 195 countries and 7 continents, we all derive from one common ancestor and share a history with one another. It is this relation that drives researchers' desire world-wide to uncover our past. Historians often begin by gathering information from ancient records and documents. However, such documents can only provide insight into the past few thousand years. Archaeology and genetic genealogy predate written records, and thus serve as a tangible resource for information about ancient civilizations. Our project combines these three aspects of historical information, written records, archaeology, and genetic genealogy, to gain insight into the history of Venetians and their supposed predecessor: the ancient Venets.

At the forefront of historical research is the use of historical documents and manuscripts. These records serve as tangible resources that can provide one of the most detailed accounts of historical events. The Venice State Archive houses an unbroken history of the city of Venice. It contains records from the various offices and branches of the Venetian government to the birth and death certificate of Venetian citizens, and almost everything in between. However, these records are becoming worn and deteriorated as a result of being over used. The Venice State Archive has currently implemented two systems to make steps toward improvement: SiASVe and Project Divenire. SiASVe is an online index though which the records within the archive can be searched. Meanwhile, Project Divenire is a series of digital reproductions of manuscripts that is available to the public online. Despite the efforts of the archive, thousands of manuscripts still remain that have no digital reproduction or information. These programs also do not allow for the collaboration of transcriptions, they simply make manuscripts more easily accessible. Records still receive more use than necessary under this process, and research is never shared among scholars, which leads to repetition. A program that has been designed by previous Worcester Polytechnic Institute teams fulfills each of these goals. This program, *uScript*, facilitates archival research through the use of a database of both records and their corresponding transcriptions. By providing a cumulative and collaborative research tool, scholars will have the ability to share their

research and transcriptions as well as to access the archival collections remotely. By increasing the amount of use of these ancient, fragile documents, we can make steps toward their preservation. Our project aimed to promote and advance this program. It involved both its promotion and its completion. A promotional website, uscript.org, was renovated to better convey the many components of *uScript*. Additionally, internships and school project opportunities were established to facilitate the completion and eventual implementation of the program. Our efforts were intended to further *uScript* by any means within our boundaries. The Venice State Archive will greatly benefit from its implementation and the many suffering manuscripts within its collection will be preserved.

Archaeology within Venice is both crucial and constant. As the modern city of Venice lies directly above the ancient city, artifacts are constantly being uncovered. An archaeologist is required to be present at every municipal excavation due to frequent amount of discoveries that occur. Floods and other natural forces pose a constant threat to the remains of archaeological sites. As Venice's topography enables new artifacts to constantly be discovered, an efficient system to record and manage archaeological data is crucial. Recently, an online program called *ArchEasy* was designed by Worcester Polytechnic Institute students to store archaeological descriptions, identifications, and locations in an online database. Although this program ultimately promotes a simpler method to manage and modify data, it does not ease data collection on-site. If artifacts are catalogued and recorded in a simpler manner, more time will be allotted to expand on the existing data. For each layer of each that is uncovered at an excavation, it is necessary to properly document any observations. Additionally, forms and descriptions are annotated for each artifact that is uncovered. The time that is spent hand-writing notes and descriptions at each excavation site is time that is not spent excavating. Time is an important factor in archaeological work, especially in a city such as Venice where water poses a constant threat to artifacts and archaeological remains. Our project focused on the data recording process within the field. We designed a mobile application, intended for Android mobile phones, that allows archaeologists to quickly and efficiently record the necessary information during excavations in a timely and organized fashion. This application will ultimately ease archaeological work by improving the speed of record keeping and enabling immediate online data storage of artifacts in a user-friendly interface. We also created a second mobile application for the public, which allows users to view various sites throughout Venice that had been

previously excavated.   These two programs provide a simpler means for obtaining archaeological data and exposing significant finds to the general public.

The city of Venice was created as a refuge for those escaping the invasion of Attila the Hun and the Germanic Lombards in the 5th century A.C.E. However, little is known about those living in the Veneto region prior to that. To fill in the gaps in history where archaeology and manuscripts fall short, we are exploring the field of genetic genealogy. Genetic genealogy combines the traditional research of genealogy with the analysis of DNA. While genetic analysis can scientifically prove relations amongst different people, the historic evidence of genealogy will show *how* they are related. The Genographic Project is currently analyzing DNA samples from all over the world to form relationships between populations and determine ancestral migratory paths. There are presently over 100 genetic samples from Venice that have been analyzed by the Genographic Project. However, Venice itself is a mixed population and no single ancestor could be determined previously from these samples; the mystery still remains. A significant amount of historical research has been compiled and evidence of their ancestors is scattered throughout Europe and northern Africa. Our project has established efforts to combine this historical research with the genetic analysis from the relevant geographical locations.   Combining genetic analysis from each location with the relevant historical history will give us insight into the relationships that these historic populations once held with those of the ancient Venets.

2. Tools for Historical Research Through Ancient Documents

A complete account of the history of Venice is documented in the 78 kilometers of shelved manuscripts in the Venice State Archives. Although these documents preserve most of Venice's history, they have a limited lifetime; many are deteriorating as a result of their age and repeated use. Scholars are constantly referencing and transcribing these manuscripts, but the process is neither efficient nor collaborative — the documents are unknowingly reused, continuing to wear. A program was developed by previous Worcester Polytechnic Institute students to transcribe and store the voluminous stock of archival information in hopes of preserving these ancient manuscripts. Although not complete, *uScript* is designed with the intent of improving the research process by creating an accumulative database of transcriptions and corresponding images of manuscripts. This program lacks the proper funding and publicity to reach its full potential. Without support, efforts for this program cannot continue and programming cannot be completed—the current method of research remains unchanged. Our project aimed to expose an alternate method to the inefficiencies of the current transcriptional process by advertising and promoting *uScript* for both public awareness and future funding. The implementation of this program would be a tremendous asset to the research process within the Venice State Archive.

2.1 Background

Written documents give one of the most accurate accounts of past events. While oral versions may become distorted with time, written records remain unchanged. They are authentic and genuine regardless of age. The following sections outline the importance of archival information and the current processes of its research.

Figure 1 Venice State Archive

2.1.1 Venice State Archive

Venice quickly became a powerful city within Europe from the time that it was founded. It was established as the center of trade in

Eastern Europe until the discovery of the passage around the Cape of Good Hope, and the Republic itself was among the most influential European powers[1]. Because of its prime location and political supremacy, Venice came in contact with a majority of European powers. This history can be seen through the diplomatic documents which are now stored in Venice's archive. The Archive of the Republic of Venice is among the largest collections of written documents, alongside the Vatican and Simancas [2]. It is "described by contemporaries as a 'treasury' and 'the heart of the state'[3]." There is no archive in Europe with equivalent historical value or richness in detail.

The Venice State Archive is currently located in the Campo dei Frari. It houses over 78 kilometers of shelving[4], containing documents and manuscripts that date over one thousand years old[5]. However, as a "live" archive, it is constantly receiving new pieces and its already vast collections are always expanding. The documents within the archive contain political records, as well as governmental and personal records including laws, bills, certificates, books, brochures, and everything in between. Each office within Venice's government generally held its own documents and records in a personal library. It is a matter of congratulation that such an incredible amount of documents are still preserved within the Venice State Archive today. For centuries the records of the Republic of Venice were stored in the personal libraries of their respective offices. It is for this reason that such an outstanding number of these records still exists in spite of the constant threats to ruin them such as floods, fires, and revolts.   It was only in 1815 that the Austrian government resolved to collect and arrange all of these documents in one single location, the convent of the Frari, which still houses this collection today. The Venice State Archive is considered a "live" archive, The preservation of these records is extremely important, as they describe and display the history of so many nations in Europe. Many of these documents are written on parchment, making them extremely susceptible to deterioration. They are under constant threat from natural forces, which have already destroyed a vast amount of ancient records. Time itself progresses their deterioration, as the collections are constantly being

[1] Brown, Horatio. *Studies in the History of Venice.* New York, NY: E.P. Dutton, 1907. Print
[2] *State Archives of Venice - Project Divenire.*Web. 10/5/2010
        <http://www.archiviodistatovenezia.it/divenire/home.htm>.
[3]De Vivo, Filippo. "Ordering the Archives of Early Modern Venice." *Archival Science* 28.3 (2004): 195-216. Print.
[4] Venice State Archive □ Archives up to the Treaty of Campoformio." *United Nations Educational, Scientific and Cultural Organization*. Web. 02 Jan. 2011. <http://portal.unesco.org/ci/en/ev.php-URL_ID=17824&URL_DO=DO_TOPIC&URL_SECTION=201.html>.
[5] Venipedia. *Venice State Archive.* Web. 10/6/2010. <*http*://venipedia.org/index.php?title=Venice_State_Archive>.

handled by the visitors of the archive. The condition of these archival pieces is still a major area of concern, as they continue to be handled and used on a regular basis.

2.1.2 Researching in the Venice State Archive

Although scholars may not be aware of it, the research process within the Venice State Archive can be both redundant in inefficient. Those wishing to access the information within the archives must locate their sources manually, and perform their research within the Study Room, seen in Figure 2, which is an expansive room available to those wishing to reference the collections of the archive. Throughout their research, scholars produce transcriptions that are purely their own and generally not shared. This leads to duplicate efforts in research and repetition of transcriptions, as scholars are accessing the same manuscripts for similar

Figure 2: Venice State Archive's Study Room

information. The general research process followed by scholars is portrayed in Figure 3.
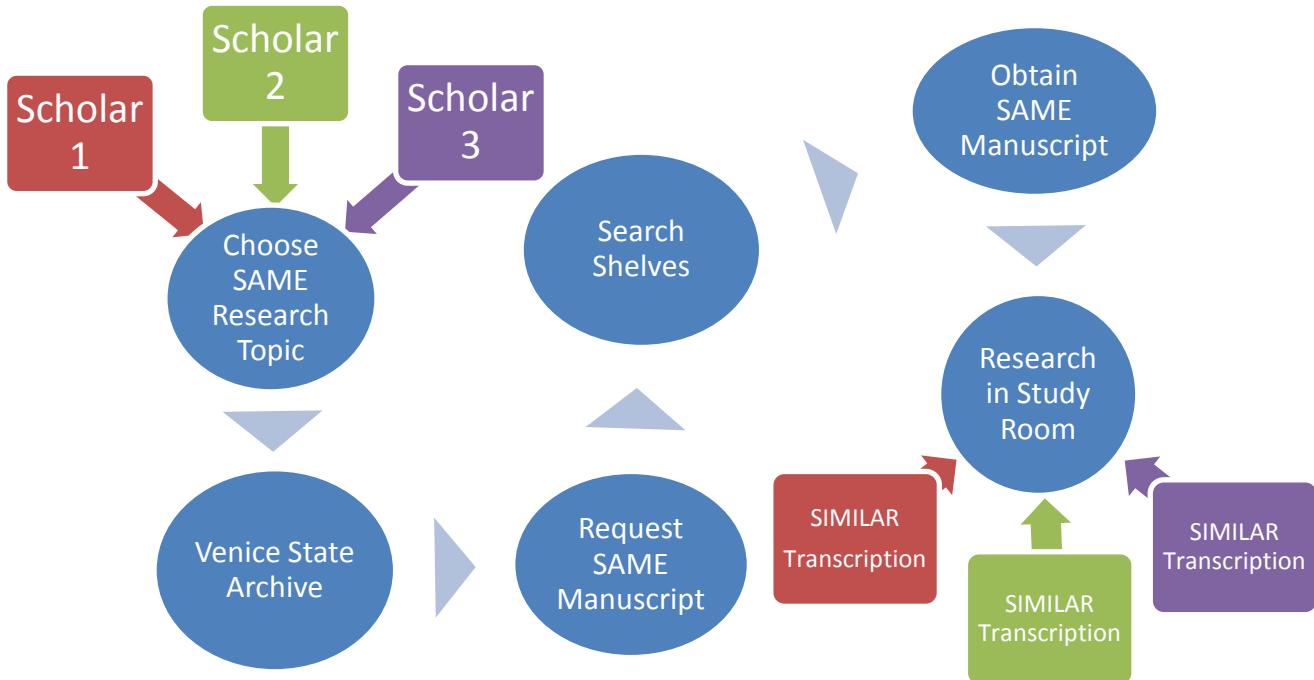
Figure 3: Current Research Process within the Venice State Archive

The redundancy that occurs within the research process progresses the wear of manuscripts considerably. The constant manipulation of these already fragile documents leaves their contents illegible and eventually ruined. Primary accounts of events are being destroyed and can never be replaced

2.1.3 Detrimental Results of Research Process on Manuscripts

The result of this repeated overuse of manuscripts within the archive is becoming both apparent and obvious. Each time one of these ancient, fragile documents is handled its contents become increasingly illegible. The aged parchment that composes these manuscripts was generally not created to withstand a vast amount of wear. Simple contributing factors such as the natural oil found on the hands of each researcher can pose threatening results, leading to the destruction of pages and ultimate illegibility of text. It is the repeated use of these records through the redundant research process

Figure 4: Evidence of the abuse that manuscripts are receiving.
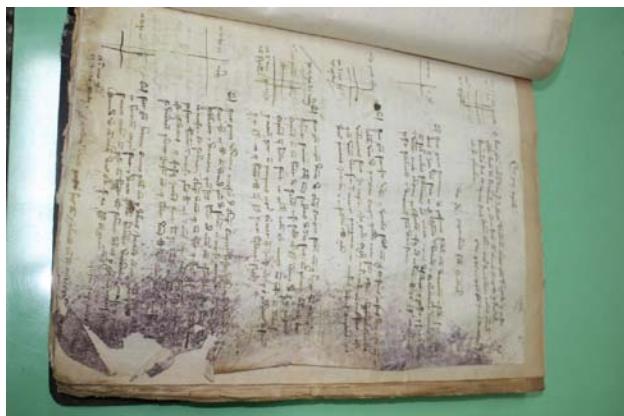
Figure 5: Deteriorating Manuscripts

explained previously that leads to their deterioration and ultimate destruction. Many records within the Venice State Archive have now been moved to a private store and made inaccessible to the public to preserve their contents and prevent further damage. The condition of these documents continues to degrade, as the current research process continues to be practiced. Although this redundancy is not intentional and often unbeknownst to those that promote it, it is apparent to those at the Venice State Archive who witness its effects every day. The need for a change is only increasing with time, as the condition of these records continues to worsen.   The valuable information contained within these written accounts is irreplaceable, and what we lose we can never regain.

2.1.4 Current Technology

The worsening condition of the collection of the Venice State Archive has inspired a number

Figure 6: Screenshot of the SiASVe Technology available online through the Venice State Archive Website

of projects aimed at improving the current system. The archive itself has introduced two systems which are both available online and open for public use. SiASVe, or Sistema Informativo dell'Archivo di Stato di Venezia (Information System of the State Archives of Venice), is one of the current online tools available in the Venice State Archive[6]. It is an online guide to the collections held within the archive. This tool organizes these

documents in a variety of ways, and is constantly being updated and expanded. It

---

[6] *State Archives of Venice - SiASVe.*Web. 10/5/2010 < http://www.archiviodistatovenezia.it/siasve/cgi-bin/pagina.pl?Lingua=en>.

gives the basic information of the recorded sources, such as the creator, a short description, and any keywords or subject terms that may help describe the specified item. The user may enter his/her desired search terms and would receive a number of suggested sources fitting the given criteria. This program is publicly available to anyone with internet access. Although this program does not contain the entire collection of historical documents within its database, it is constantly being expanded and added to by the achival experts within the Venice State Archive.

The Project Divenire[7] is another effort put in place by the Venice State Archive to reduce the usage of manuscripts. It is a collection of digital reproductions of manuscripts that is available to the public. This is an internet-based program, accessible worldwide, through which you can view the physical pages of a number of manuscripts   For those pieces entered into this database, individual pictures of

Figure 7: Scanning Manuscripts

each page of the record can be viewed. A brief description of the document is also provided, although these descriptions vary based on the type of document. As with the SiASVe program, the Project Divenire is not fully complete due to the vast amount of material within the Venice State Archive.
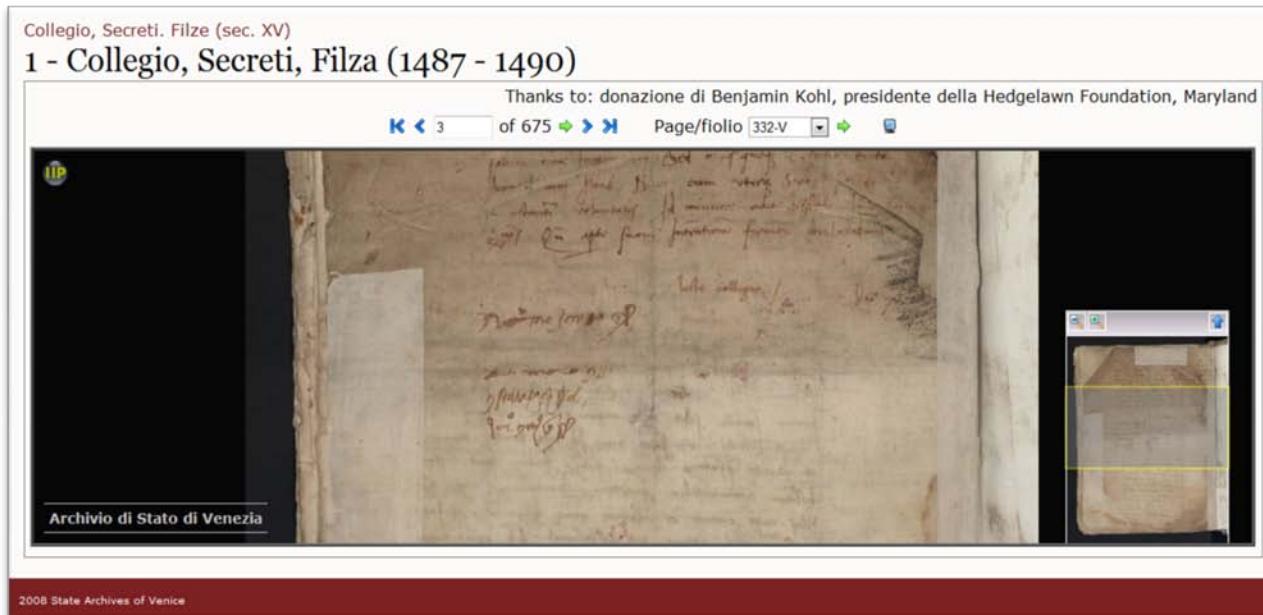
Figure 8: Screenshot of the Project Divenire, available online throught the Venice State Archive website.

It is an ongoing database that is constantly updated with new pieces. Despite their incompletion, these programs effectively take steps toward the ultimate goal of preserving these invaluable materials before they are ruined and lost forever.

---

[7] *State Archives of Venice – Project Divenire*. Web. 10/5/2010.
<http://www.archiviodistatovenezia.it/index.php?id=63&L=3>.

2.2 uScript: The Archival Transcription Assistant

In the 1990's a program was developed to facilitate research within the Venice State Archive. This program, *uScript*, was designed to assist in the painstaking task of transcribing historical documents and manuscripts. *uScript* is an electronic tool which contains digital reproduction of manuscripts as well as their corresponding trasncriptions. It provides a searchable database that is both collaborative and cumulative, which will improve the research process in the future and decrease the excessive amount of usage that ancient documents are currently recieving.

2.2.1 Transcription Assistant

The main feature of *uScript* is the Transcription Assistant. The transcription assistant is responsible for the correlation of transcriptions with the relevant text in the manuscript image. Upon selecting a manuscript page from within the program, the user will be brought to a split-screen window, such as seen below in Figure 9. The user may then transcribe the corresponding document at their discretion; they may either transcribe the entirety of the document or only a

portion of it.
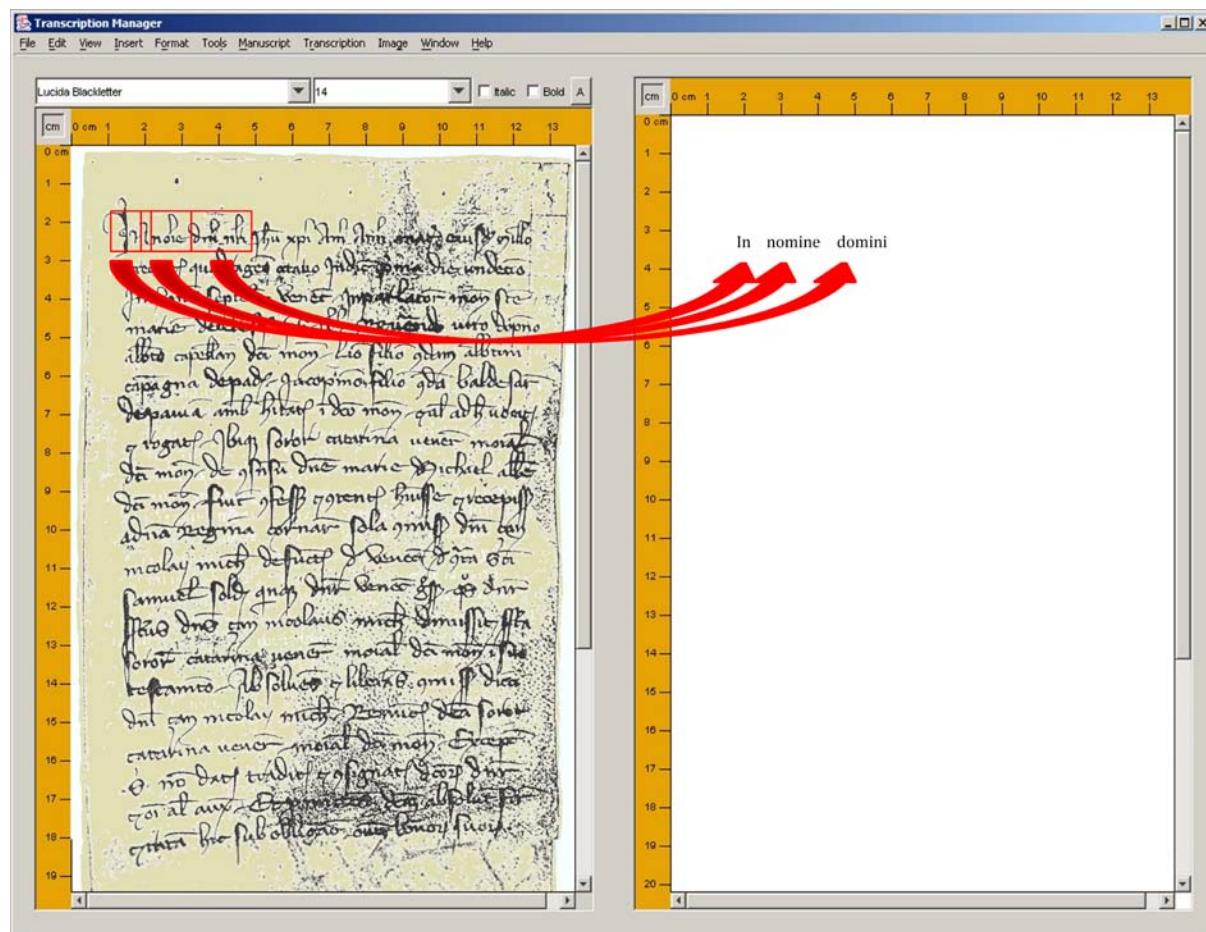
Figure 9: Example of the uScript Program

As seen in the above figure, each word that is transcribed into the database is correlated with its respective word in the manuscript reproduction using a series of wordboxes. This one-to-one wordbox-to-text correlation will allow for handwriting recognition capabilites in future software. When the user feels he is finished transcribing, his transcription will be stored in the *uScript* database, whether complete or unfinished, along with the image of the relevant manuscript. Future researchers wishing to access this manuscript would the not only recieve the image of the manuscript, but also the transcription that was previously created for it. They will have the ability to correct, edit, and add to previously existing transcriptions as they see fit. This collaborative and cumulative aspect is a key component of *uScript*. The existence of such a database will allow for transcriptions to be produced more quickly and accurately.

2.2.2 Archival Assistant

The Archival Assistant is the *uScript* component responsible for searching and organizing information within the database. The current organizational system of the archives is not always ideal for researchers. Documents were organized as they were created, not by location or date, and filed through a series of indexes[8]. Using *uScript*, the researcher will enter a series of search criteria based on the topic he wishes to investigate. Based on these criteria, the Archival Assistant will locate various manuscript pages that correspond to the desired topic. The pages will be presented to the user as a list of options, upon which he will choose the one that he feels best fits his peronsal research. After a specific page (or pages) are chosen, the Transcription Assistant tool would be used to enter a corresponsing transcription.

2.2.3 Contribution Accountant

The Contribution Accountant is the final contributor to the *uScript* program. After the transcription is entered, it is rated by the Contribution Accountant based on the number of previous entries the user has transcribed as well as the accuracy of their transcription. This feature was designed to keep *uScript* as accurate and precise as possible so that it may provide a tangible resource for future scholars.

2.2.4 Missing Resources for Completion
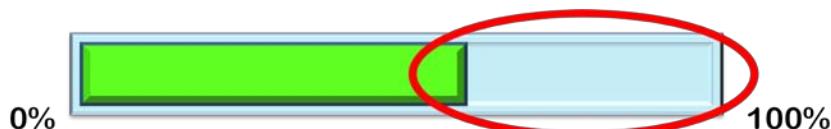
Figure 10: uScript is not complete.

Each of these three componenets plays an important role in the overall transcription and research process. The collaborative and cumulative aspects of the program will be beneficial to the archive and its collections. In recent years, the students of Worcester Polytechnic Institute (WPI) in the United States have taken interest in the facets of *uScript* and its use iwthin the archive. However, despite recent efforts, the program is not yet implemented within the Venice State Archive and its benefits remain unknown to the majority of the public. The program does not currently possess enough financial support to continue with its production. Without proper funding, additional improvements cannot be made and the program remains incomplete.

---

[8] http://www.springerlink.com/content/e2q4t86j542tv771/fulltext.pdf

Additionally, support must be located for the completion of the software and internal coding of the program. Our project aimed to facilitate these advancements because we felt they held priority in the overall implementation process. The importance of *uScript* within the archive is obvious, and progress must be made before its components can be put into use.

2.3 Methodology

Our project continued the efforts of previous years to further the *uScript program* in order to improve both manuscript preservation and the processes by which they are studied. Our team focused on promoting *uScript* to establish means for its completion and future implementation. We redesigned an informational website to promote the program, as discussed in section 4.2.1. We also established opportunities for summer interns and school project teams to improve the program's code, described in section 4.2.2. As a means to promote and complete *uScript*, we requested monetary aid from potential sources of funding, outlined in section 4.2.3.

2.3.1 Promoting uScript

Our team is promoting the implementation and advancement of *uScript* through the use of an informational website. Having a concrete and central location for information will provide our present and future collaborators with an easily accessible reference that is dedicated exclusively to *uScript* and its many features and specifications.

The *uScript* website was improved to more accurately convey the program's intentions and specifications. The 2009 Worcester Polytechnic Institute Origins Research Team created the current website using WordPress and templates from the previous efforts of WPI students to promote *uScript* and gain support for its development. However, the website had become inactive and was no longer available online. Our group restored the *uScript* website and made it once again available to the public. Upon restoration, the website and each of its components were evaluated for accuracy of content, functionality, and ease of use. We found that the website contained an abundance of useful information, but needed reorganizing and repairing. Many of the internal links were no longer working or available, and required restoration as well. It is important for future users and collaborators to have the ability to access as much information as possible when visiting the *uScript* website. The goal of this website is to promote the program

and gain support for its advancement, so it is mandatory that its contents be accessible.

Figure 11: Screenshot of Previous uScript Website

The content of the website pages was another area of concentration. Descriptions of the program and its components were previously described by the former project team. However, we felt that it was best to revise and edit these descriptions, and expand on them where necessary.

To improve the existing website, we first familiarized ourselves with the old website. Each page of the website was thoroughly examined and evaluated prior to any changes being made. The content itself was analyzed and links and images were tested for functionality and relevance. The website contains fifteen published pages, along with internal links to articles and publications relevant to *uScript*. Among these pages was the grant proposal written by Fabio Carerra to the National Endowment for the Humanities, written to gain support of the program and fund its completion. Within this document is the complete details of the *uScript* program as well as the hypothesized future additions that would be made. Our team based the content and details of this website off of this grant proposal, as it was the most tangible explanation available to us.

The first portion of the website to be changed was the "uScript Software" section. This page gave a one sentence description of uScript and had two links at the bottom, titled "uScript Demo" and "uScript Components." The "uScript Demo" was a previously existing page that was designed to load a demo of the program upon selection. However, the link was broken and the demo could not be located. The "uScript Components" link led to page with a small description of the three components and three links

containing more lengthy descriptions. We decided that this information should be more easily accessible from the website directly. The information on the "uScript Software" page was replaced by a concise but accurate description of the program and its three components. Additionally, the "uScript Components" page was deleted and three pages with detailed descriptions of the three components were added to the website. These pages are easily accessible from the "uScript Software" page or from the main home screen.

The next section of the website to be revised was the "Get Involved" section. This section consisted of five pages, the overall "Get Involved" page, as well as supporting pages titled Crowdsourcing, Intellectual Property Concerns, Collaborative Filtering, and Reputation Management. This section of the website was designed to explain how the common user can get involved in transcribing manuscripts on uScript, how their work will be ranked within the website, and how their intellectual property will be conserved. Revisions on this portion of the website included redirecting the aforementioned pages to make them more accessible. This was completed through the use of a navigational drop-down tab, located at the top of the page. Additionally, the content within the page was edited to display accurate and current information regarding the program.

The "About" section of the website was the next portion that was edited. This section is the first description of uScript that a visitor to the website would see. It is therefore crucial that this information be up-to-date, as well as concise and clear. The content within the "About" section was heavily edited, evidence by the fact that the "Purpose" and "Vision" pages were completed rewritten. The page on the website to be revised was the "Funding" page. This page was also completed rewritten, since the information had to be changed and updated.
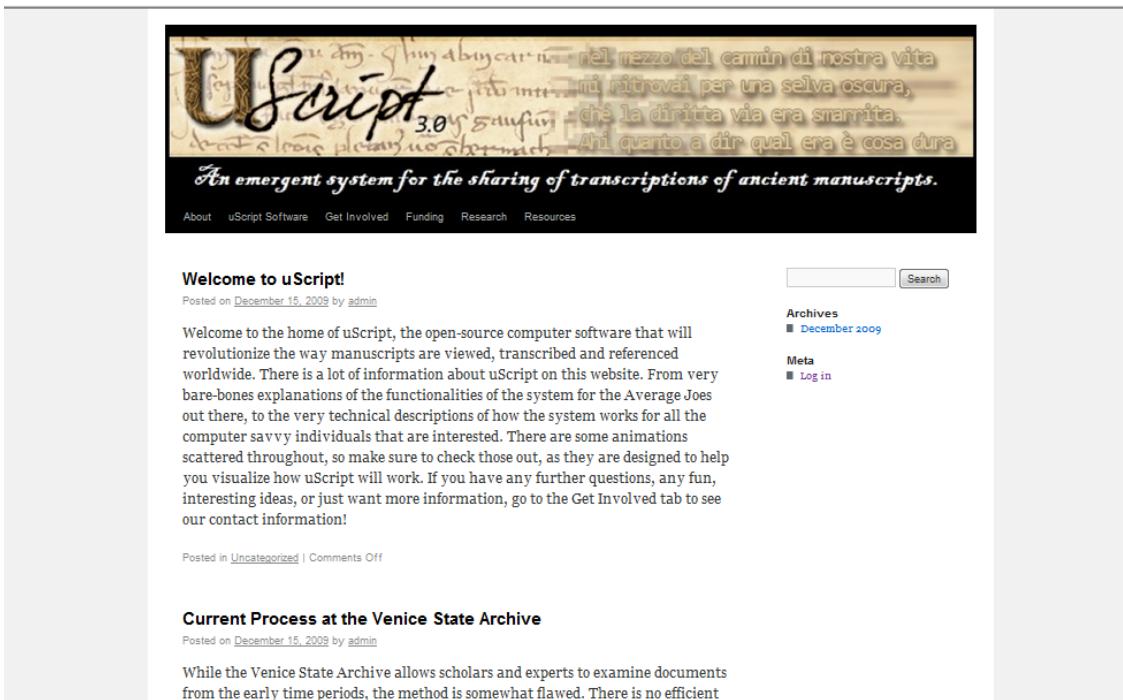
Figure 12: Screenshot of New uScript Website

The final changes to the website were aesthetic changes. These include changing the background theme from "Elegant Grunge" to "Twenty Ten". The background theme was changed so the website would have a cleaner, more inviting look. Additionally, the navigation bars on the "Twenty Ten" theme can be located and used more easily.

2.3.2 Advancing uScript

This portion of the project focused on improving and advancing the *uScript* code. The code is currently unfinished and therefore the program is incomplete. To improve the software and functionality of uScript, we are setting up school projects and summer internships, which we hope will facilitate its advancement. These programs are intended for students studying Computer Science or Interactive Media and Game Development, as they would possess the most useful background knowledge for this task. We have set up student-internship opportunities thought WPI's Career Development Center. This center promotes internship opportunities to all students within WPI. Our internship is designed to take place in Venice, and in collaboration with the Venice State Archive. The student would be compensated for room and board and live in Venice, Italy for the duration of the internship.

Figure 13: WPI's Career Development Center

Students completing their degree at Worcester Polytechnic Institute are required to complete a Major Qualifying Project within their major area of research as a graduation requirement. We established a Major Qualifying Project that focuses on improving the uScript code and making it open source once it's completed. This project would be completed abroad, in Venice, and the cost of room and board would be added to tuition. This is designed only for those Major Qualifying Projects within the Computer Science and Interactive Game and Media Development fields, as the format of coding is complex and requires extensive background knowledge. Although there is previously existing code for numerous components of *uScript*, it is written in a number of different coding languages (i.e. HTML, C++, Java, etc.) and would be difficult to understand for those who are under qualified.

Both of these opportunities aim to advance and eventually complete the *uScript* program's code. A considerable amount of work has been done in the other areas of *uScript*, and completing the code is the primary task that must be addressed before the project can proceed further.

2.3.3 Funding uScript

The implementation of the *uScript* program cannot continue without financial support. Funding needs to be located prior to its implementation and usage within the Venice State Archive. Our team has making efforts to locate such funding. Funding request letters were written, as seen in Appendix A, and sent to various relevant organizations and individuals. By acquiring financial support, we are hope to use the money to fund the previously mentioned student internships and projects or any external programming work that will advance and complete *uScript*. A grant proposal, as seen in Appendix B, has also been created in hopes to obtain the necessary support from various organizations. Also, if *uScript* can be made into an open source program, enabling the public to improve and expand the code, then it will be eligible

for Google Summer of Code grant money. The money collected by these efforts will enable uScript's public accessibility and ultimately the implementation of the program.

2.4 Results

Our project ultimately set up the advancement and implementation of uScript, as we personally do not possess the knowledge and expertise to continue with the coding of the program. The improved website, described in section 4.3.1., is an informational resource for all collaborators in the future. The student projects and internship opportunities that were established are outlined in section 4.3.2.

2.4.1 *uScript* Website

As the public becomes more aware of the potential that *uScript* has within the research process at the Venice State Archive, it will naturally gain technical support and financial aid. Within the archive, an increasing number of ancient manuscripts are being put away every year. These manuscripts have been deteriorated over time due to the extreme amount of repetitive usage that they receive. Those wishing to access these manuscripts within the archive will soon be denied of their request.

It is imperative that *uScript* be implemented to avoid this. The goal of a public website containing information specific to *uScript* is to promote and advertise its existence. There are several areas in the development process where external support and resources are needed. As stated previously, before this project can continue in the future, it must obtain funding. The resources to expand the code and complete the program are currently insufficient. This is the issue that our website aims to solve. By making information about *uScript* publicly available, we hope to one day gain the support that is necessary to implement this program with the archives and facilitate future historical research.

The newest edition of the uScript Website, located at www.uscript.org, has been completely revised, redited, and reconstructed. As explained previously, the content of the website was almost entirely changed to convey more accurate and easily-understandable descriptions of the program and relevant components.   Many of the links within the website led to destinations that no longer existed, and thus needed to be redirected or replaced. The website can now be seen in the new "Twenty-Ten" skin, with a navigation bar that cleaner and more user-friendly. Our aim is

to gain public support of the program and its importance through the use of this website. Having an accurate and easy-to-navigate website is a key aspect in spreading public knowledge and interest.

2.4.2 Advertised Projects and Internships

The programming within *uScript* is currently in a variety of states, and a variety of styles. There is not a rubric of any nature that previous programmers were told to follow while creating this transcriptional tool. Although each portion of *uScript* has worked previously, they have not all worked *together*. This is due to the styles that they have been written in. Each section of program, or code, is written in a language (i.e. HTML, C++, Java, etc.) that was specifically chosen by the creator based on his/her preference. The future students working on this project will need to be able to decipher through this pre-existing code, and create a program that is functional as a whole. Worcester Polytechnic Institute is has a variety of majors within the computer programming fields. It is for this reason that we felt it was most beneficial to advertise such an opportunity within the school. The completion of the code would be a substantial advancement toward *uScript*'s implementation. The details of future goals and intentions of the *uScript* programming can be found in Dr. Fabio Carerra's grant proposal in Appedix B.

2.5 Conclusion

The completion of *uScript* will improve the overall efficiency of research within the Venice State Archive and will assist in preserving the current collections. The deterioration of manuscripts is of extreme concern within the archive. Continuous use of the documents decreases their longevity, each use worsening their overall state. The physical condition of these documents is as far from optimal as is the process of their research. Not only does a scholar spend time waiting to receive one of the thousands of documents, he or she is also prohibited from leaving the archive due to each document's fragility. Because no collaborative study is involved, the transcription process is redundant, individualistic, and damaging.

Worcester Polytechnic Institute's contribution of *uScript* presents a new opportunity to for change. The implementation of assistive transcriptional technology will slow document wear and improve the current process. *uScript's* ability to search manuscripts by content, assist transcription, and rate individual contribution will result in a more efficient method of manuscript

study. Transcription will be accumulative, collaborative, and assisted. Although this program promises more effective and efficient methods of scholarly research, it is not yet complete.

In order to progress towards the employment of *uScript* within the Venice State Archive, more resources for funding and labor are necessary. Financial support will enable monetary compensation to be allocated for student interns or professionals completing programming the *uScript* code. Funding letters, such as the request letter exampled in Appendix A, or a grant proposal is recommended to be sent to relevant organizations of interest in order to obtain monetary aid. A grant proposal may provide opportunities for a more efficient implementation of the program. One specific organization that should be targeted as a potential source of support is Google's Google Summer of Code. A grant from Google Summer of Code will result in not only funding but an open-source project status for *uScript.* If *uScript* becomes an open-source effort, it will be made available for online collaborative programming. With increased contributions to the code from a variety of professionals, this program can be completed and implemented sooner.

Due to *uScript's* early stages of development, acquiring public interest may be challenging. However, appeal to those unfamiliar with the program's benefits is crucial for obtaining the necessary resources for its completion. One method that our group used to publicly advertise *uScript* is the improvement of its informational website, available at www.uscript.org. By further promoting the program and its benefits through the website, interested individuals or collaborators will be exposed to a descriptive yet concise source of information about *uScript* without additional efforts to request information from the program's developers. By continuous public promotion, *uScript* will continue to gain public interest and will progress in the direction of ultimate public use.

Although *uScript's* development is not yet complete, the compilation of its database information can begin as soon as possible. According to advice from an administrator of the Venice State Archive, the program's database can begin to be populated with numerous manuscripts that have already been transcribed in entirety. Transcriptions and corresponding scans of the manuscripts are available within the Venice State Archive. By transferring these files into the accumulative database of *uScript*, foundations for content recognition and transcription assistance will be established. Interested users will be able to explore the new functions that this program has to offer even before programming is fully completed. An individual will be able to view currently transcribed manuscripts to conduct historical research without physically being

within the Venice State Archive. Additionally, if an individual begins to transcribe a new or existing manuscript, the program will be capable of assisting the transcription by generating suggested words that it may recognize from the already stored transcriptions and corresponding images. If *uScript's* database is populated with existing material now, the process of collaborative transcription can begin and will result in an earlier accumulation of scholarly information.

Overall, the implementation of *uScript* into the Venice State Archive will provide scholars with an accumulative, collaborative, and assistive tool for historical research and manuscript transcription. Being an online program, *uScript* will enable individuals to study in the comfort of their own homes yet collaborate to create shared public information. With a continuously increasing word bank and collection of documents, this program will ultimately recognize words and create transcriptions without extensive input from scholars. The implementation of *uScript* relies on the accumulation of resources: the more funding and promotion *uScript* receives, the more contributors it will attract. In turn, the more contributors *uScript* obtains, the more manuscripts will be transcribed and the more historical accounts will be available for public study. An increased amount of public information will result in new historical discoveries and, ultimately, an even clearer account of the past.

3. Tools for Facilitating Archaeological Processes

Although archaeology is one of the most reliable sources of historical information in Venice, the processes to facilitate on-site research are inefficient. Archaeologists create maps to track the locations of their finds; however, these maps quickly become outdated as new artifacts are discovered. Floods and other natural forces pose a constant threat to the remains of archaeological sites. Additionally, the current procedures for recording new information are both tedious and time consuming. We aimed to increase the efficiency of the current data recording process by designing and implementing an Android mobile application. Additionally, through the creation of a second mobile application, we have made information regarding previously excavated archaeological sites available to the public.

3.1 Background

Archaeology is a field of study that uses the physical remains of humans and past civilizations to try to understand how past peoples lived. Archaeology combines science, historical documents, and artifacts to analyze past cultures. Because archaeology is one of the most important tools Venetians have for exploring and understanding their history, an efficient method to record and store data is imperative.

3.1.1 The Study of Ancient Peoples

It is stated that archaeology is the "systematic study of past human life and culture by the recovery and examination of remaining material evidence."[9] This illustrates that archaeology is not only involves the study of artifacts, such as pottery and carved tools, but also the study of the significance behind these artifacts. Since 99 percent of human history occurred in prehistoric cultures, archaeology is one of the few means we have to discover and understand the majority of human existence.[10] Although these cultures left no written records, they did leave behind a plethora of evidence which include relics, such as graves, buildings, tools, and pottery. Even though archaeology is the main tool used to study and interpret prehistoric cultures, is also useful as a tool to study more recent history.

---

[9] http://www.answers.com/topic/archaeology
[10] Renfrew and Bahn (2004 [1991]:13)

Archaeology is generally thought to be the study of prehistoric civilizations. However, the field is much more expansive and can include relics from any era in history, including the 20th century. For this reason, archaeology is not one succinct discipline but has many sub-disciplines which focus on specific aspects or time frames.

Prehistoric archaeology focuses on civilizations that were in existence before written records were created. The Veneti is an example of such a civilization

Figure 14: Veneti Burial Urns

that is considered "prehistoric."[11] The prehistoric sub-discipline puts emphasis on relics such as pottery, rock tools, and graves; where correlations and conclusions can be draw from patterns in the ceramics, in the rock carvings, and in burial style. An example of this can be seen in the burial urns which are characteristic of the Veneti and are often found in locations where the Veneti are theorized to have lived.

Conversely, the main concentration of historical archaeology is on civilizations that existed after the formation of written documents.[12] In this sub-discipline, historians use archaeological evidence in correlation with written documents to acquire an understanding of the culture of the given civilization. This is the type of archaeology that pertains to Venice. In addition to the relics and artifacts found within the city, Venice is home to one of the largest collection of historical written documents in Europe. It is due to this multitude of information that historians are able to gain such insight into the history of Venice.

3.1.2 Archaeology in the Venetian Lagoon

One of the many mysteries that still captivate Venetians today is the origin of their city; and, how an inhabitable, marshy lagoon was transformed into one of the richest, most powerful nations

---

[11] *Prehistoric Archaeology (archaeology).*Web. 10/17/2010
  <http://www.britannica.com/EBchecked/topic/474764/prehistoric-archaeology>.
[12] *Historical Archaeology.*Web. 10/17/2010
  <http://archaeology.about.com/od/historical/Historical_Archaeology.htm>.

of the Mediterranean. It is evident that this information is invaluable to the city, since the Venetian government mandates that an archaeologist be present at every municipal excavation, due to the excess of artifacts that lie underneath Venice.[13] One instance of the extraordinary relics uncovered in the region, is the recent finding of the ancient city of Altinum.
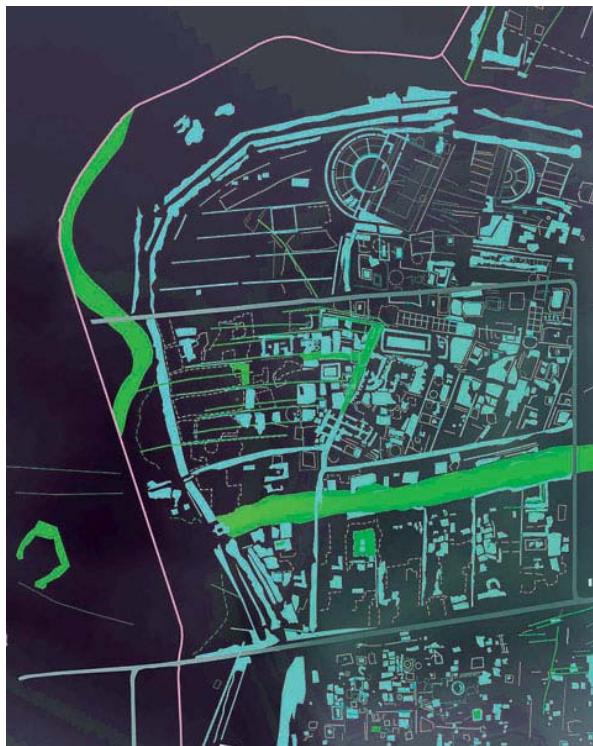
Figure 15: Infrared Photograph of the Roman city of Altinum

Altinum is an ancient Roman city which was one of the most "important trading and seafaring [centers] of the Adriatic".[14] It is theorized that during the mid-5th century the city succumbed to Attila the Hun and its inhabitants escaped further into the Venetian Lagoon, subsequently leading to the creation of the modern-day city of Venice.[15] Although historians had written proof of the city's existence and location, they had no evidence that could illustrate the structural layout of the city.

Nonetheless, a project leader with the Archaeological Institute of America, Andrea Ninfo, was able to photograph elements such as walls, foundations, and a canal with the use of visible light and near-infrared aerial cameras.[16] This astonishing discovery has given historians and archaeologists a picture of the city and a key to learn more about the history of the region. This has given the inhabitants of the Veneto a piece of their past and a link to those who once inhabited the area.

---

[13] 2008 IQP
[14] Chris Irvine. "Ancestor City of Venice Unearthed." *Telegraph* 17 Sep 2008 2008. Web
[15] Chris Irvine. "Ancestor City of Venice Unearthed." *Telegraph* 17 Sep 2008 2008. Web
[16] *Roman Venice Discovered.*Web. 10/5/2010 <http://www.archaeology.org/0911/trenches/roman_venice.html>.

3.1.3 Current Procedure for the Documentation of Archaeology Data

Venice is a city with a long, complex history that spans over a millennium. For this reason, archaeologists are continuously digging and searching for artifacts and information in the lagoon. An archaeologist's work is both difficult and meticulous, and the paperwork involved is cumbersome. There are a number of forms that require manual completion before leaving the excavation site. The Soprintendenza Archeologica per Il Veneto, the Archaeological Superintendent for the Veneto, requires a large amount of information, including measurements, pictures, and drawings. As a result, the process for recording data within the field is long and tedious.
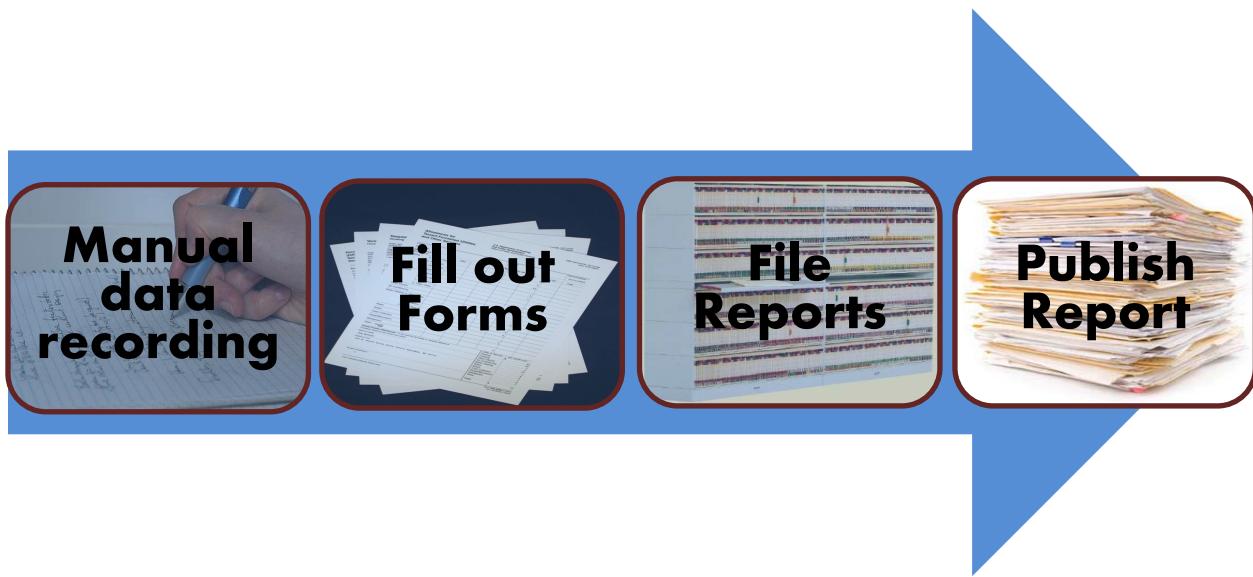
Figure 16: Archaeological Process

The first step in this process involves recording date while on site during an archaeological excavation. This generally includes the locality and civil number, the director of the excavation, the company carrying out the excavation, the chief of the building site, and a number of other requirements necessary to prepare for an archaeological dig.[17] After this, the projected dig site must be photographed in detail. These photographs include: the entrance of the dig site, the general area and contents of the site, panoramic views, and photographs with metric references. Every photograph must be taken with a minimum of three exposures to account for any defective photographs and thus prevent discrepancies in the information collected. The archaeologist must then develop a plan of action for the area of the excavation,

Figure 17: Scheda Sceleto

Figure 18: Scheda Organica

and communicate the information to the leader of the site while abiding by any terms that the leader has set.[18] Additionally, any relevant descriptions regarding the site must be recorded.

Archaeological sites are generally excavated by layers. Archaeologists will uncover a layer of the earth, stopping at their discretion, and evaluate any findings and observations. For each layer, the same data recording process must be followed; pictures and descriptions must be noted as well as any relationships that the layer may have to its surrounding (i.e. cut by, covered by, or filled by another layer or object). In addition to these observations, archaeologists are responsible for filling out the appropriate forms that correspond to their

---

[17] *Archaeology - Venipedia.*Web. 10/5/2010 <http://venipedia.org/index.php?title=Archaeology>.
[18] PG1 of Dr. Marco Bortoletto's Manuale di Procedura,02/27/2007

findings.   These forms include: the Scheda di Intervento Archeologico, the Unita Stratigrafica form, the Indice di Foto, the Scheda Sceleto, and the Scheda Organica. Each form is used to catalog and document a different component of the archaeological dig. The Scheda di Intervento

Figure 19: Indice di Foto

Archeologico, the archaeological project form, comprises all of the administrative information necessary for the excavation. The Unita Stratigrafica is a form that must be filled out for every layer of earth that is uncovered. Once the form is completed it is assigned a number, which then corresponds to that specific layer of earth. The next form filled out is the Indice di Foto, which is an index of every photograph taking during the excavation. The subsequent forms, the Scheda Sceleto and Scheda Organica, are specific to the archaeological dig and the artifacts discovered during the excavation. The Scheda Sceleto is a form which records skeletal findings, while the Scheda Organica is records organic findings. Once the archaeologist has finished cataloging all the required information an expert from the Soprintendenza Archaeologica must fill out an RA form for every layer and artifact uncovered during the dig.[19]

After the site has been excavated and the appropriate descriptions and observations have been recorded, the remainder of the work is completed off-site. The information that was gathered at the excavation site must be organized and compiled into a report, and filed away accordingly. A report is filed for each site that is excavated, even those that are simple and mundane. Of the reports filed, only a small portion are released to the public and studied by other individuals.

As demonstrated, the procedure for documenting archaeological data in Venice is both complex and tedious. The information that must be recorded while on site is expansive and can take valuable time away from the physical excavation process. This limits efficiency while on site, especially where there is not always an abundance of time to begin with.

---

[19] *Archaeology - Venipedia.*Web. 10/5/2010 <http://venipedia.org/index.php?title=Archaeology>.

40

3.2 Methodology

Our team is facilitating archaeological work by creating an efficient process for on-site data recording. Based on the current needs of archaeologists in Venice, we are designing a mobile application compatible with Android smart phones. The application will ease the recording, storage, and transfer of data. The application user needs are outlined in section 2.2.1, the design process is explained in section 2.2.2, and the promotion of the application is covered in section 2.2.3. In order to publicly promote this application, we are creating a secondary mobile application, described in 3.2.4., that draws potential public appeal for archaeology-related smart phone applications.

3.2.1 Understanding Specifications for Application

Certain procedures are necessary to be followed in order to successfully examine a dig site and its components. Because every component of an excavation has to be photographed, analyzed, and documented, specific information is collected to ensure that the data is both standardized and thorough. Documentation of a site's various aspects is a lengthy and somewhat inefficient process. A large number of photographs are taken during the dig, and numerous paper forms are required to be filled out both on and off site.

Some dig sites are analyzed over a period of days and involve several archaeologists. Since the only method for one archaeologist to continue examining a site where the previous archaeologist left off and to avoid repeating prior analysis, he or she must refer to existing documentation of that site's components. Due to the amount of paperwork that such an archaeologist must look through to understand what to analyze next, time is often wasted or duplicate paperwork is produced. A method of quickly storing basic information, and then transferring it from one individual to the next, would ease archaeological work on-site. Based on specifications provided to us by modern archaeologists in Venice, we are designing a mobile smart phone application that can be used to document essential information about an excavation in a simple and user-friendly manner. The application, once ultimately coded, will be publicly accessible and compatible with Android mobile devices.

The function of the smart phone application is to store basic data about an archaeological site and its components to facilitate further work and analysis. As specified by archaeologists in Venice, information about the site, layers within the site, and artifacts within the layers must be systematically recorded. The location of the site and its specific code, assigned by the

archaeologist, must be stored. Before beginning digging, pictures are taken to document the initial state of the site and surroundings. Each picture is identified with a code consisting of a sequenced number (e.g. 1, 2...n) following the site's code. After the initial state of the covered site is surveyed, digging begins in a distinct procedure.

A dig site is uncovered in layers. Each distinct layer, distinguished by composition, age, or professional discretion of the archaeologist, is numbered. The highest, or ground, level is assigned a number (e.g. 100); the level below the ground level is a number one higher than the ground level (e.g. 101), and every subsequent layer below that level is numbered correspondingly (e.g. 102, 103...n). In cases when an archaeologist finds significant material above the initial ground layer (layer 100) a number one less than that of the ground layer (e.g. 99) is assigned to this higher layer and smaller numbers are assigned to every new layer found above those. After adding a new numbered layer to the application, a brief description of the layer must be recorded in the simplest manner possible—either manually or via a voice-recording function of the application. Additionally, pictures of every layer must be taken and stored before beginning to remove a new layer of the site.

In order to substantially document an excavation, relations between individual components of the dig site—layers and the artifacts found within them—must be established. Every layer of an excavation is related to at least one other layer that comes within its direct contact. A layer is recorded as "cut by," "filled by," or "covered by" another existing layer or artifact. When an artifact is found, an archaeologist must document its characteristics. The application has to store photographs of the artifact, record a brief description, and classify the object. According to archaeologists, an artifact is categorized as either organic or inorganic. Organic objects are classified as bone, carbon, or shell; inorganic objects are either masonry, ceramic, metal, glass, coins, mortar, or plastic. After documenting the object's type, the archaeologist has to relate its position to existing layers or artifacts as mentioned above.

By recording and storing basic information about an excavated site using a mobile smart phone application, an archaeologist will be able to relay new data to another individual more efficiently. Although completing required paperwork to sufficiently catalog a dig site cannot be avoided, the initial process of documenting basic data will be simplified with this mobile application. Because the input information will be immediately stored in an external electronic database, archaeologists will be able to access and modify their collected data online.

Although this application will need coding to become available for future public use, it is currently in the design phase. Several mock-ups are created and final designs are being examined to promote the application and enable future coding.

2.2.2 Designing the Application with Photoshop

Based on archaeologists' need for a more efficient method of storing data at a dig site, we are designing a smart phone application which will be available for public use and compatible with Android technology. Mock versions of the application are designed using Photoshop CS5 Device Central software. Graphic templates within Device Central enable the design of the application to reflect both appearance and functions that would be characteristic of an Android mobile application. For our project we decided to use the Desire by HTC phone template. This phone has a screen size of 480 x 800px which is common on many other Android devices so that the design can also be easily displayed on phones such as the Nexus One by Google, the Droid Incredible by HTC, and the Evo 4G by HTC depending on which model people are more comfortable using. We also downloaded the Android GUI PSD v.2.0 which is a Photoshop file containing numerous elements that we used for the design of our application[20].

The following functions from the Photoshop element file were used to design our application. We first used the basic status bar that is on all Android phones including the icons for 3G, network signal strength, battery life, and the time. This was set as a base layer for all the screens, other than camera mode, to display. The next piece we used was the Tab selection. We changed this around a little to include three tabs, one for Sites, Layers, and Artifacts. Our tabs also have three different color options based on what information the user has already entered into the program. For instance a user on the Layer Tab has already been prompted to select a Site but they have not yet gotten to the Artifact Tab. In this case, the Site Tab will be black showing that it has already been completed, the Layer Tab will be white indicating that this is the tab currently in use, and the Artifact Tab will be grey signifying that no information has been entered for this item yet. As an added feature once you move on to a tab to the right the entered information from the previous tab will be displayed on the Tabs up top, such as the site and layer numbers.

Another design we included was the dropdown menu buttons. This aspect was used for each of the select options of site, layer, and artifact. Once clicking on this button a swipe motion scroll bar box will pop-up. In Photoshop we created this affect by overlaying different groups of

---

[20]Android GUI PSD version 2.0. Web. 12/11/10 <http://www.vgrishin.me/android_gui.html>.

the content and each time moving the words up a space. The scroll bar box will contain, depending on the tab the user is on, either a list of all the sites entered as well as their addresses, a list of all the layers, or a list of all artifacts found. The site scroll will also have an icon indicating whether an artifact was found on that site. The same icon will also appear on the layer scroll along with an icon indicating the type of material the site consists of. These numbers will be written either in bold font to indicate a complete object, or regular font to show an ongoing item. If the site, layer, or artifact does not show up in the list the user also has the option of adding a new one simply by clicking the appropriate button or plus sign.

The layer and artifact tabs contain a Gallery view of pictures the user has taken for each object. The user can easily scroll through these by clicking the arrows on either side of the photo thumbnails. Photos can be added by clicking on the Camera Icon underneath the gallery. When doing so the user is brought to their phone's camera mode and can take as many pictures as they desire. The next object the user can enter is the Description on both the layer and artifact tabs. This is done by clicking the white box to the right which then displays a virtual keyboard and large text box. Here the user can manually type in a description of what they see, or if the user's mobile phone is equipped with voice-recognition software they can simply speak into their phone which will then transcribe everything they say easily. A short excerpt of the description is displayed on the main page once completed.

On the layer tab the user now can enter Relations by clicking the plus sign. This will display dual swipe motion scroll bar boxes. The box on the left will display the relation choices of covered by, cut by, and filled by, while the box on the right will list all of the layers that have already been entered into the database. Once the user selects this it will be added to the main screen under the Relation title. The final part of the layer tab involves check-boxes to indicate the type of Material the layer is made out of whether it is sand, clay, or silt.

The artifact tab also contains an Artifact Type selection. This involves radio buttons, which are circles where only one at a time can be selected. The two main options are Organic and Inorganic. Once either one of these is selected a swipe motion scroll bar box will also be displayed with further classification options. The organic option contains the three subcategories of bones, carbon-based, and shells. The inorganic alternative contains the seven subtypes of masonry, ceramics, metals, glass, mortar, coins, and plastics. When a selection of one of these is made the information will be displayed on the main screen as well as an edit icon which allows

the user to go back if they made a mistake. These are the basic design elements involved in creating the mobile application in Photoshop. A more detailed explanation of the application and its functions is provided in section in 3.3.1 of the results.

3.2.3 Creating a Second Application

The second application is created using the Layar Reality Browser. This application shows the user the world around them by overlaying real time data on top of the real world, as seen through their mobile phone's camera[21]. Layer uses a combination of GPS, the mobile phone's camera, and a compass to pinpoint the user's exact location and field of view, using this information the Layar application then overlays selected data onto the camera screen. In order to use Layar the team needs to have two accounts on the site itself, one for publishing and one for developing. In addition to the Layar access we also need access to Bluehost, which hosts most of the web parts of the Venice project site. With both of these and instructions from the Layar wiki as well as another Venice Project Center student, we are able to create a database.

We produced an excel sheet with basic information, that we obtained from the archaeologists, about six archaeological dig sites around Venice. These sites will serve as a test to show the possibilities of the Layar application. The information collected for each site consists of a name, GPS location, picture, and three lines of text. We uploaded this data onto Bluehost in the form of a POI table. This allowed the team to go back to the Layar website and create our own layer with the given data points. On Layar we are now able to test the device online as seen in Figure 11 as well as on phones to fix any issues that we may encounter. Once the testing phase is complete, we need to request publication approval. This will allow our layer to be used by anyone in Venice with the Layar App already installed on their phone. We hope that this application will inform visitors about the history of Venice as well as promote future advancements in archaeological technology. Please refer to section 3.3.2 which provides a more detailed explanation of the application and its functions.

---

[21] What is Layar?. Web. 11/29/10 <http://site.layar.com/download/layar/>.
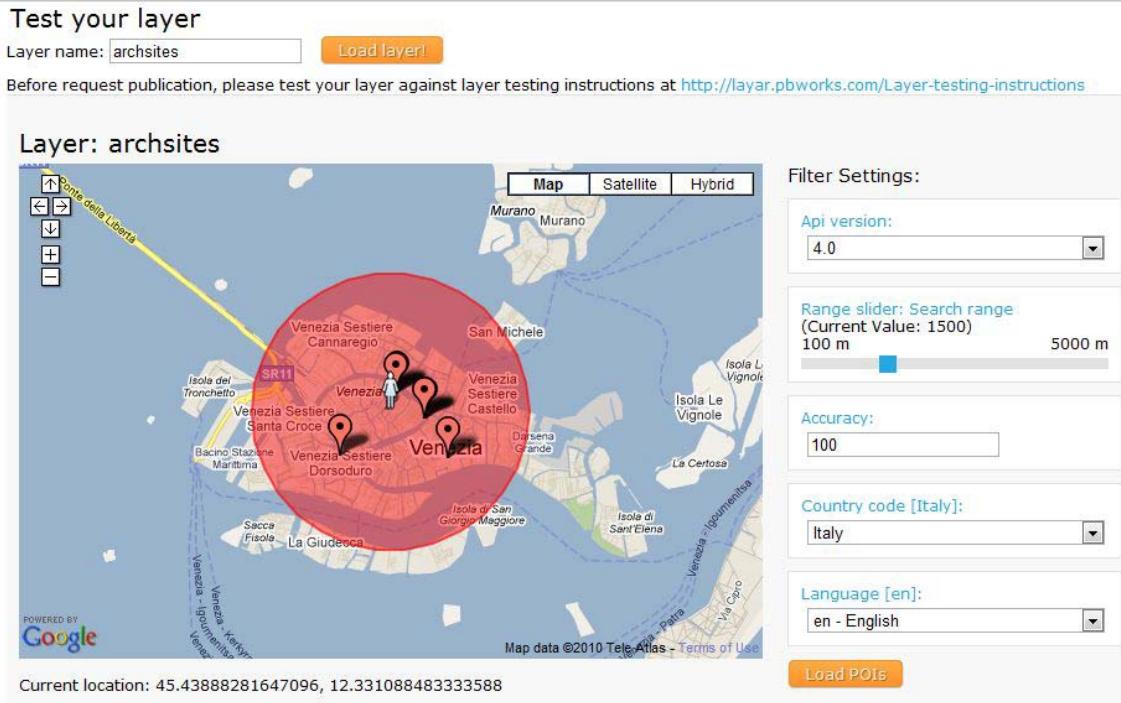
Figure 20: Layar Application Test Screen

3.3 Results

Our project ultimately facilitates archaeological data recording during excavations, as described in section 3.3.1, and provides information on previously excavated archaeological sites to the public, as described in section 3.3.2.

3.3.1 Archaeology Mobile Application

We have designed an archaeological mobile application that will facilitate archaeologists' field work. The application, ArchApp, is focused on the recording and data collection process that archaeologists undergo in the field. It consists of a number of screens, through which the archaeologist can input data, such as the layers of earth they have excavated as well as artifacts found within these layers. Although ultimately paper documentation must be filed for each archaeological dig, this tool will shorten the amount of time spent recording data at the site, as well as assist in organizing information for each site during excavations. This application will also interface with an online database where information can be accessed and uploaded from both the mobile phone and desktop computers.

Each excavation site is uncovered by archaeologists in layers. Every layer of earth has relationships with its surroundings that must be recorded. Any artifacts found within each layer

are also recorded. Every detail of an excavation needs to be recorded thoroughly, so as best to illustrate the site and process when referenced in the future. Our application is designed based on the information we gathered regarding the archaeological excavation process.

The application opens to the main screen seen in Figure 12. The top of this screen contains three tabs for easy navigation. The first tab, which is active at startup, is the Site Tab, next is the Layer Tab followed by the Artifact Tab. From this screen, the user may choose from a list of existing sites which have been previously entered into the database, or create a new site if they do not find their choice amongst those listed. The scroll bar that pops up when selecting a site contains basic information such as the site number, address, and an icon to depict whether any artifacts have been found there. If the site information is in bold font this depicts that the site has been completed, while normal font indicates that the site is still active.

Figure 21: Site Tab of Archaeology Application

Figure 22: Layer Tab of Archaeological Application

After choosing a site, the user will be brought to the Layer Tab as seen in Figure 13 where they will be prompted to select a previously entered layer or create an additional layer. For each layer, the user will be able to take photographs, record a description, add any relationships with surrounding layers, and select the material of the layer. The user will be able to take photos by clicking on the camera icon which will bring the user to their phones camera mode. Once completed with taking photos, the user will be brought back to the Layer Tab where the photos taken will be displayed in a gallery view. This allows for easy scanning and movement through all the pictures taken at the given layer. The user can then add a description of the layer either using the virtual keyboard on their phone or voice recognition software, if installed, to automatically transcribe what the user is saying. The next step will be to add any relations that this layer has with others. This is done using a scroll bar menu, as before with one for choosing the type of relation, either covered by, filled by, or cut by

and the second choosing another layer number. The last step on this screen involves selecting what kind of material, either sand, clay, or silt, the layer is made up of.

Figure 23: Artifact Tab Archaeological Application

Once all this information has been added you will be brought to the Artifact Tab as seen in Figure 14 where you will once again be prompted with a scroll menu, this time of Artifacts on the specific site and layer you are working on. You also have the option to add a new Artifact by simply clicking the plus sign at the top of the screen. In the Artifact Tab you are given the same options for taking photographs and recording a description as in the Layer Tab. This tab however asks you to enter what type of artifact you are dealing with. First you must select either the Organic or Inorganic circle. Either choice will provide the user with a scroll bar selection. The organic button has three subcategories of Bones, Carbon-Based, and Shells. The inorganic button contains the seven options of Masonry, Ceramics, Metals, Glass, Mortar, Coins, and Plastics. Once you make your selection of artifact type the type will show up on the screen. However, if a mistake was made there is also a small edit icon next to Artifact Type that will bring you back to the choosing process.

Once you are done entering all the information on the Artifact Tab you can choose to enter additional artifacts, go back to the Layer Tab, or even go to a different Site using the Site Tab. For each site an infinite number of layers and artifacts can be added. All the information is recorded chronologically and numbered as appropriate. The application automatically stores the information the user entered in an online database, which is easily accessible from a computer so that the reports can be written in a quicker manner. The implementation of our design of this application would greatly improve the current archaeological process therefore increasing overall efficiency. It would simplify the overall data collection process, which is a key factor in archaeological research today.

3.3.2 LAYAR Mobile Application

The Layar application on your smart phone displays points of interest over what your camera view normally shows of the world around you as seen in Figure 15. These points are circles of varying sizes that depending on your proximity to their location become larger the closer you are. As you move your phone around the circles also move and new ones will show up in

Figure 24: Main Page Layar Application

different locations. All of these points are clickable and upon that give you more information about each point as seen in Figure 16. In the archaeology dig sites Layar that we created, you will

Figure 26: Point Information Layar Application

be given the name of the dig site, a picture of the site as it was during excavation, and a few lines of text. This text includes the archaeologists on site, the managing editor, and a brief description of what was found. You can also view these points in Map mode which using Google maps places markers at all the archaeological sites and also shows your current location. There is also List

mode which gives a list version of all the points on the application including the picture and text.

The application itself also has a take me there button which will give the user directions using Google Maps on how to get to the dig site if you wished to see, in person, what it looks like now. This Layar application has the ability to include many more dig sites in Venice once the reports have been made public. However this is just a preliminary tool to draw attention to

Figure 25: Map View Layar Application

the many technological advances out there that can be useful in the field of Archeology. Hopefully this Layar application will spark the interest of archaeologists and others interested in Venice so that future groups can expand upon the accessibility and wealth of information found at these dig sites.

3.4 Conclusions and Recommendations

A multitude of archaeological reports are produced every year in Venice due to the city's potential for abundant historical finds. When significant archaeological material is uncovered, its documented information is published and made available to interested scholars. Unfortunately, not only are much of these finds never exposed to the general public, each excavation is further concealed from public attention by being restored to its uncovered state. Our mobile smart phone application presents an opportunity for change: visuals and descriptions about important archaeological findings at given locations can be viewed through a user-friendly interface on a phone's camera screen.

To further the implementation of this mobile application, additional enhancements should be made. Although little descriptive information is currently imbedded into the generated displays of each archaeological site, the base software for the application, Layar, limits the amount of information the application can display. Because there is not enough available space for a more thorough description of each site, a URL address that links to a website should be embedded into each display of archaeological data. Corresponding website pages containing a more elaborate depiction of the sites should be written, preferably through the Venice Project Center's *Venipedia* site.   Thus, to view a more thorough description of a previously unearthed finding, an individual will be able to access an informational website by selecting a link on each displayed screen.

In order to produce an archaeological report, archaeologists must document every component of an excavation using standard forms and procedures. Although the completion of some documents cannot be bypassed during the production of these reports, the inefficient process for manually recording archaeological data on-site can be substituted with a simpler one. Our designed smart phone application will transform the process of manually hand-writing data in a limited amount of time into more efficient means for data documentation and storage.

Although the functions of this application meet the specifications of most Venetian archaeologists, further improvements to the design can be made. As previously described in section 3.3.1., the design of the application contains three tabs and three corresponding screens

which prompt the user to input photographs and descriptions of sites, layers, or artifacts, respectively. According to professional critique of archaeologists, the measure of each layer's elevation should additionally be prompted for on the "layer" screen. A layer's elevation must be recorded as the perpendicular distance from the layer's surface to a specific reference point designated by the archaeologist. This reference point, or "monument" point, should be noted within the "layer" screen for future reference. Besides prompting a user for data input, these three component-specific screens also display information relevant to each excavation component (site, layer, or artifact). One improvement that can be made to the application is the addition of a displayed map of the site's location. This map should be an automatically generated map of the site's exact location and should be displayed on the "site" screen to provide the user with a readily available visual of a dig's geographical location.

In order to develop, code, and ultimately implement this data recording application into archaeological practice, sources of funding should be sought. A funding request letter or a grant proposal and a use case can be sent to relevant organizations that may potentially be interested in the furthering of the application. By obtaining funding, monetary compensation can be provided to future student interns or professionals who are willing to develop and implement the application.

The development and ultimate implementation of this archaeological data recording application promise benefits not only to Venetian archaeology, but similarly to archaeological practices worldwide. Although certain details of required documented data may differ from one practice to the next, the overall excavation process is procedurally similar. By obtaining feedback from various archaeologists around the world, a universal version of an archaeological application can be developed. Such an application will enable archaeologists, regardless of location, to bypass the slow and inefficient processes of manually documenting excavation data by providing a new tool for efficiency. Not only will the input information be recorded in an organized and user-friendly interface, it will be easily modified off-site through an online database. Overall, the employment of a smart phone application would both enhance and simplify common archaeological practice.

4. Tools for Genetic Research to Determine Origins of the Ancient Venets

Genetic genealogy is a recently developed field in historical research, combining genetics with traditional genealogy. In other words, researchers collect and analyze deoxyribonucleic acid (DNA) samples to determine genealogical relationships between populations. Because this study focuses on precise genetic patterns which have been passed down from one generation to the next, scientists can trace the migratory paths and ancestral lineages of our predecessors. National Geographic and IBM are currently promoting the Genographic Project as a means to determine the origins of populations across the globe. Despite National Genographic's effort to complete DNA analyses in most regions of the world, little is known about the genealogy of Venice's founding population. The origin of the Veneti remains a mystery. Theories to explain the origin of the city's original inhabitants do exist, however. Since archaeology and written accounts have failed to prove any one of these theories, more precise methods such as genetic genealogy have been undertaken to identify the origin of the Veneti. Although some DNA samples have been analyzed, there is a lack of genetic evidence to prove any definite relations between the Veneti and other ancient populations. The insufficient number of valid samples allows for theories to continue to be made, but does not provide enough information to confirm them. To determine both the origins of the Veneti and the history of the people of Venice, our project intends to continue sampling and analyzing DNA from relevant locations in collaboration with the National Genographic Project.

4.1 Background

Currently, the only available information about the ancient Venets can be obtained from archaeological finds and historical records. Although the combination of these two sources of information provides some thorough accounts of the history of the Veneto and the eventual foundation of Venice, not enough evidence has been uncovered to create a complete depiction of this ancient population. Even less is known about the origins of the ancient Venets. Where did the predecessors of Venice's ancient founders migrate from?

Clues to this population's origins are found throughout Europe. Evidence which potentially connects Veneto's ancient residents to other groups of people is geographically dispersed in patterns which are hypothesized to be the regions of Venet's origin. Although theories about their past exist, archaeology and history alone cannot prove any one of the theories. A more precise

science, genetic genealogy, can fill in the gaps that artifacts and written accounts leave out to ultimately establish a connection between the ancient Venets and their migrant founders.

The following sections describe the history of the Veneto and the current evidence pertaining to its founders' origins, the application of genetic genealogy to establish connections between relevant populations of the past, and modern genetic genealogy studies of interest which may contribute determining the Venets' true origin.

4.1.1 Early History of the Veneto

The earliest accounts of the ancient Veneto date back to around 2,000 B.C.E., when the Euganei people first populated northern Italy, settling in the Alpine foothills, the Raeti, and the Trentino and Alto Adige valleys[22]. Their control over north-eastern Italy was short lived, however. According to legend, a mysterious Adriatic settlement drove the Euganei westward around 1,000 B.C.E., and continued to populate the Veneto for the next 300 years. Around 700 B.C.E., Greeks and Etruscans discovered the region and soon merged with the Adriatic

Figure 27: The Veneto Region

inhabitants; these three ethnic populations soon became known as the Veneti, and their homeland, the Veneto[23].

Numerous invasions reshaped the history and societies of the Veneto. Soon after, the Romans invaded and conquered the Veneti, amending the new territory to the Holy Roman Empire. Later in the fifth century, Alaric the Goth and Attila and the Huns began their invasions throughout Italy[24]. Because numerous cities were destroyed and villages pillaged, devastated residents sought

---

Image courtesy of http://z.about.com

[22] *Euganei (people).* http://www.britannica.com/EBchecked/topic/195032/Euganei

[23] Butterfield, Bruce J. *Livy's History of Rome.* 2006. http://mcadams.posc.mu.edu/txt/ah/Livy/Livy01.html

[24] Ammerman, Albert J. *Memoirs of the American Academy in Rome.* 2003. http://www.jstor.org/pss/4238806

Figure 28: The Hun Invasion

refuge in Veneto's lagoons. Although the Huns failed their conquest and eventually left, many of the escapees remained in what are now the cities of Grado, Torcello, and Venice. In 586 C.E., Italy was invaded by the Germanic Lombards. Similarly, many Romans that fled to the Veneto remained in its lagoons even after peace ensued. It is thus believed that the modern city of Venice was first established by refugees in combination with the ethnic residents of the lagoon[25].

4.1.2 Origins Theories

The history of the Veneto has welcomed various populations, many of whose origins are yet unknown. Who were the Adriatic people that drove out the Euganei and settled north-eastern Italy? Where exactly did the lagoon's refugee populations come from?

Although a more concrete science is needed to prove the true origins of the Venets, historical research and archaeological finds yield various evidences which may connect this ancient population to others throughout Europe. Up to date, three major theories hypothesize the migratory origin of the ancient Venets: the Paphlagonia Theory, the Lusatia Theory, and the Brittany-Wales theory. Additionally, various finds expose information that may link the Venets to their ancient population of origin.

4.1.2.1 The Paphlagonia Theory

The oldest and most documented, the Paphlagonia Theory states that the Adriatic population which drove out the Euganei and later fused with the Greeks and Etruscans migrated from Paphlagonia, or modern-day northern Turkey. According to historical research, the Paphlagonians fled their homeland after the Trojan War, moving through the Adriatic and eventually discovering

Figure 29: The Paphlagonia Theory

---

[25] *History of ITALY.* http://www.historyworld.net/wrldhis/PlainTextHistories.asp?gtrack=pthc&ParagraphID=glq

and inhabiting the Veneto.

The beliefs held by the Paphlagonia Theory are, for the most part, evidenced by ancient documents. In fact, literary pieces from both Livy and Homer are the basis of its ideas. The largest mentions of Paphlagonian emigrants can be found in Livy's historical accounts. Living in Italy during the reign of Agustus, as well as events such as the Trojan Wars, Livy kept records of any changes throughout the ancient empire[26]. According to Livy, a revolution during the Trojan War resulting in the death of the Enetae leader, King Pylaemenes, forced the Enetians to flee Paphlagonia and migrate west. Livy's accounts indicate that the Enetians moved along the Adriatic coastline and, after overthrowing the Euganie, ultimately settled in what is now the Veneto[27].

Book 2 of Homer's *Illiad* also mentions the Enetae. Like Livy's historical documentations, the *Illiad* documents the Enetae leaving Paphlagonia after the death of their king during the Trojan War and migrating west to eventually settle in the lagoons of the Veneto[28].

4.1.2.2 The Lusatia Theory

Less supported than the Paphlagonia Theory, the Lusatia Theory is based on research compiled in the 20th century. The theory first surfaced when Slovene author Jozko Savli attempted

Figure 30: The Lusatia Theory

to explain European history by tracing major European populations to the Venets, a people that originated in Lusatia in north-eastern Europe. In 1984, Savli published *Veneti: First Builders of European Community* to explain his theory. According to the book, not only did the Slovenes originate from the Venets rather than the Slavs, but most current-day European nationalities can be traced to Veneti origins. Savli claimed that the Venets' ancestors of populations in Brittany, Italy, and Austria, as well as other countries, migrated from Lusatia—modern-day Poland, Germany, and Czech Republic[29].

---

[26] Butterfield, Bruce J. *Livy's History of Rome*. 2006. http://mcadams.posc.mu.edu/txt/ah/Livy/Livy01.html

[27] (*Venetoimage - Canova e Possagno*. http://www.venetoimage.com/svc.html

[28] *The Classic Internet Archive | The Iliad by Homer*. http://classics.mit.edu/Homer/iliad.2.ii.html

[29] Pantelić, Stjepan. *Veneti Were Autochthonous in Carinthia*. http://www.korenine.si/zborniki/zbornik01/htm/pantelic.htm

Both radical and slightly counterintuitive, the Lusatia Theory held little support throughout Europe. Although the theory was often criticized by scholars, it gave rise to questions and opened possibilities for more research. Soon following the theory's first publications, historian Stephan Pantelic argued his support for Savli's ideas by documenting that the Slavs, as well as other Europeans, are linked to the Venets of Lusatia in north-central Europe. Pantelic rejected the Paphlagonia Theory. He explained that the Venets originated near the Danube River in central Europe, not in Paphlagonia. Additionally, Pantelic argues that a native population of these Venets resided in Carinthia, Austria; during the Trojan Wars, they traveled to Paphlagonia to ally with the Enetians. After the wars, the Carinthian Venets left Paphlagonia, spreading through Western Europe, the Adriatic, and northern Italy[30].

4.1.2.3 The Brittany-Wales Theory

The least supported of the three theories—the Brittany Theory—claims that the ancient Venets migrated to the lagoons of the Veneto from north-western France. The founder of the theory was Strabo, an ancient Greek historian who rejected Livy's Paphlagonia theory. Strabo

argued that the inhabitants of Gaul, or modern-day Brittany, who were coincidentally called the "Veneti" as well, spread throughout Europe and eventually settled in the Veneto. Since many Venets of Gaul migrated north to Wales, recent historical findings also evidence that this same populated may have first

Figure 31: The Brittany Theory

migrated to the modern United Kingdom, thus connecting the unknown Venets of Wales to those of both Gaul and later, the Veneto. It is this migratory group that may have later became Venice's first population[31].

Evidence of Venet Influence in Europe

[30] Steinacher, Roland. *Studien zur vandalischen Geschichte*. June 2002. http://homepage.uibk.ac.at/~c61705/DISSERTATION-Volltext.pdf
[31] *Strabo: The Geography*. July 18, 2009. http://penelope.uchicago.edu/Thayer/E/Roman/Texts/Strabo/home.html

Although there are three major theories that attempt to establish the true origin of the ancient Venets, various other evidence of this population's influence has been found throughout Europe. After years of research, scholars and archaeologists have uncovered ancient clues that may connect the Venets to other populations. Besides written accounts which state previously developed theories about Venet origins, archaeological finds, mythology, traditions, and other historical information expose numerous significant locations that may have been characteristic of the ancient Venets.

The discovery of Oetzi, an ancient man preserved by ice in the Tyrolean Alps, presents another piece of archaeological evidence suggesting a potential migratory connection of the ancient Venets' ancestors to another European population. Oetzi, named after the Oetz Valley in the Alps by his discoverers, was dated to be 5,300 to 5,100 years old. Analysis of Oetzi's location within the Alps suggests that there may have an ancient, specifically a Late Neolithic, migration from an area west of the Alps to one on the east. After substantial genetic analysis, Oetzi was found to be related to ancient populations residing in regions which are now modern France. The Veneto's location directly east of the Alps where Oetzi was found has led many archaeologist and historians to believe that Oetzi may be related to the ancient residents of the Veneto, the Venets. Although this particular find supports the hypothesis that the Venets migrated from Brittany, various other evidence reveals significant connections of the Venets to several distinct ancient populations throughout Europe[32].

The following information pertaining to evidence of Venet migration through Europe has been obtained from various collaborating scholars who have kindly supplied us with past and current research. A spreadsheet of collaborators participating in the efforts to study the Venets is available in Appendix I.

Various archaeological finds have presented new information revealing insight about the migration of the Venets throughout ancient Europe. Burial urns, used by the ancient Venets to cremate their dead, have been found across distinct regions in Europe. In some cases, the type of ceramic and the characteristic physical features of the urns found in the Veneto resemble those of

---

[32] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.

urns found at several locations throughout Europe. The geographic distribution of these urns is seen on the map below.

Figure 32: Geographical Locations of Burial Urns

Most burial urns found in archaeological sites in the Veneto have also been discovered prominently in the regions of the United Kingdom, Brittany in France, Slovenia, Slovakia, Lusatia (intersection of Poland, Germany, and the Czech Republic), and northern Turkey.

Likewise, worship of deities considered identical to those that the ancient Venets have worshiped is evidenced in a similar geographical distribution to other archaeological findings, such as the previously mentioned burial urns. As shown in the map below, goddesses such as the Venet goddess Raetia were prominent religious figures in United Kingdom's northern Wales, Brittany, Poland, Slovenia and the Veneto.
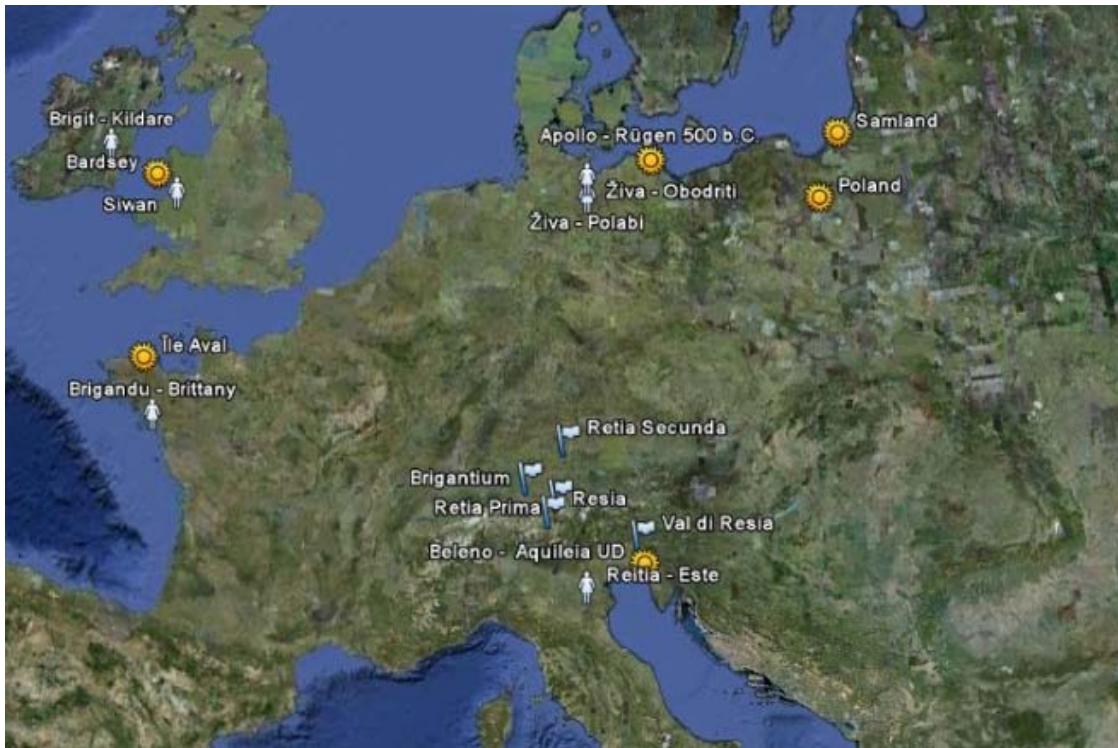
Figure 33: Geographical Locations of Similar Deity Worship

Similar patterns of Venets' evidence distributions were found upon thorough analysis of various sources of research. Historical accounts of traditions such as ceremonies involving bonfires link the ancient Venets to other ancient populations with the same cultural aspects. Many of these characteristic cultural practices are described in documents relative to not only the Venets, but cultures across Europe. Additionally, ancient Venetic inscriptions found on archaeological artifacts of the Venets have also been discovered in more northern regions of Europe such as Poland, Samland, and Germany. Amber, another evidence of the Venet's past connections to other populations, has been found in the Veneto as well as northern Europe and Brittany. Although amber may simply suggest the existence of the amber trade routes through each of these regions of Europe, it also exposes a possible migratory connection of the Venets to other groups. These and other types of archeological and historical evidences, although different, are distributed in approximately similar regions as displayed in the map below.
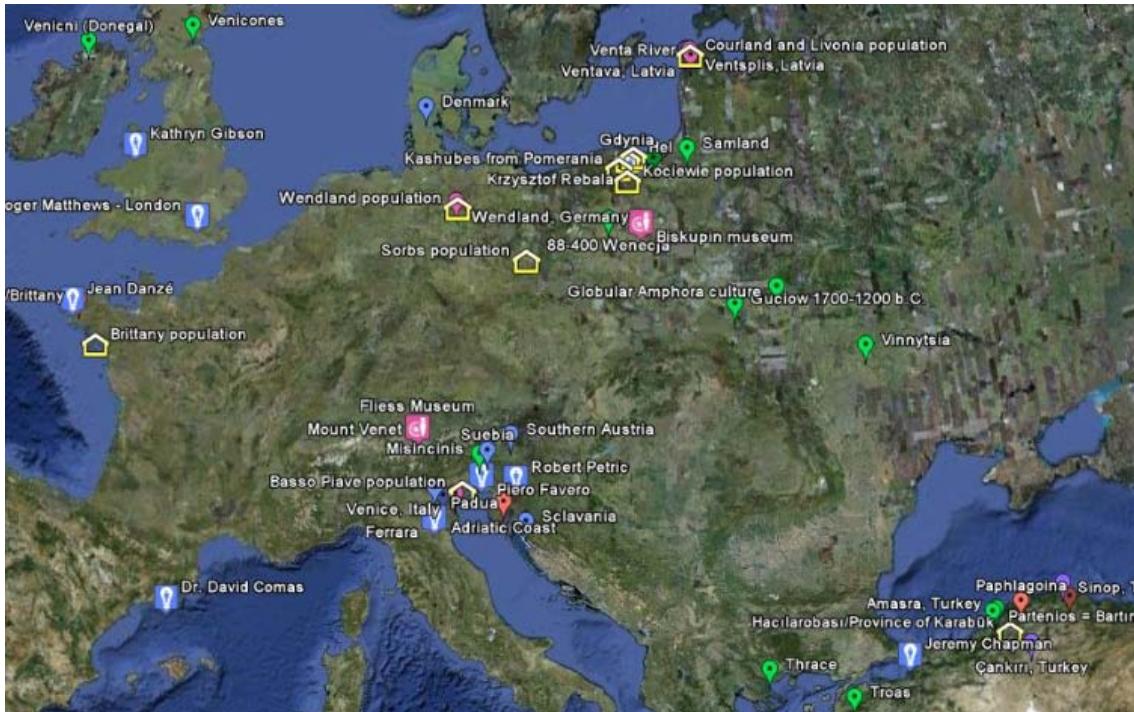
Figure 34: Geographical Locations of Other Significant Evidence to Research of Venets

These patterns of nearly identical geographical distribution of each variety of evidence imply that the ancient Venets are somehow connected to one or more of the ancient populations once residing in these regions. Substantial discoveries lead scholars to believe that the origin of the ancient Venets lies in one of the mentioned locations.

Although the major theories and the evidences that expose potential connections to the Venets are supported by substantial historical documentation or archaeological finds, no concrete resource of historical information has been found to prove their true origin. For a more tangible method of determining Venet origins, recently developed scientific studies should be applied.

4.1.3 Genetic Genealogy

Genetic genealogy is a modern science that combines genetic analysis with traditional genealogy, or the study of family ancestry. For decades, genealogists have been examining family lineages through historical records and oral traditions. Although familial lineages have been established by written accounts and stories alone, genetics can improve the study by both broadening the scope of analysis and validating the ancestral relationships found. In fact, genetic genealogy enables scientists to determine familial lineages as far back as 1,000 years[33]. In this

---

[33] Smolenyak, Megan Smolenyak. *Trace Your Roots with DNA*. United States of America: Rodale, Inc., 2004

study, samples of deoxyribonucleic acid (DNA) are examined in order to find genetic patterns that have been passed down through generations. By analyzing and comparing these patterns in various international populations, scientists can establish connections and determine the migratory paths of our ancestors by the most accurate means available today.

4.1.3.1 DNA and Genetic Recombination
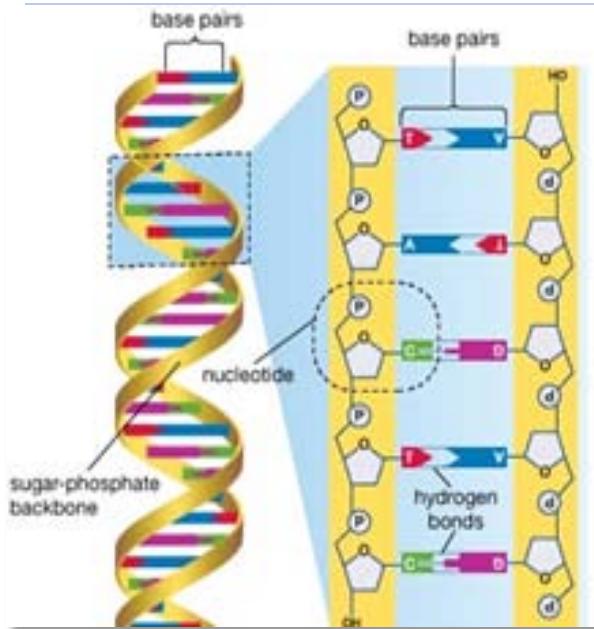
Figure 35: Structure of DNA

DNA is composed of chemical base units called nucleotides. Each nucleotide is a combination of a sugar molecule (deoxyribose), a phosphate molecule, and one of four chemical bases. These bases are purines, adenine (A) or guanine (G), or pyrimidines, cytosine (C) or thymine (T). DNA is a long double-helix shaped molecule in which pairs of nucleotide bases are held together by phosphate bonds[34]. Specifically, adenine (A) pairs with thymine (T), and cytosine (C) pairs with guanine (G). It is the order of these nucleotide bases that defines the products made from each gene. Specific sequences of three of these bases, A, T, C, or G, code for one amino acid; in turn, combinations of amino acids make up proteins, the building blocks of life[35]. Because one DNA strand acts as a template for the synthesis of the complementary strand during cell division, the sequences of nucleotide pairs and their corresponding genes are copied and passed down from cell to cell, and eventually, from generation to generation.

Combinations of nucleotides make up genes which are organized into 23

Figure 36: Nuclear DNA Components

---

[34] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.
[35] *A basic Introduction to the Science UnderlyingNCBI Resources*, A Science Primer, NIH. http://www.ncbi.nlm.nih.gov/About/primer/genetics_genome.html
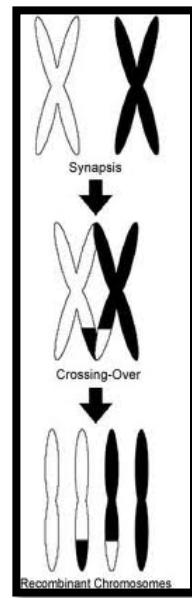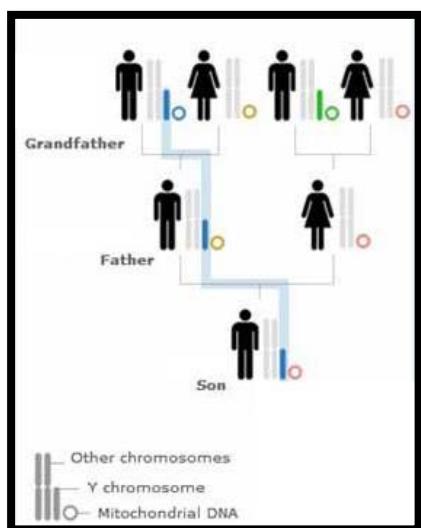
chromosomes, or structures composed of segments of DNA wound around proteins called histones. These 23 chromosomes make up nuclear DNA found in every living cell of the human body. Half of the human genome is inherited from the father, and half from the mother. In general, the genetic information that is passed down differs from the original DNA of both parents. The process which differentiates parental DNA from that which is inherited by offspring occurs during the division of sex cells, or meiotic cell division[36]. When sex cells divide, segments of DNA within a specific chromosome shuffle and recombine with segments of DNA in its corresponding complementary chromosome in a particular pair. As a result, new genetic sequences are produced, and slightly variable genome halves from both the mother and the father combine during fertilization to produce a genetically different offspring.

4.1.3.2 Exceptions to Gene Recombination

Not all genes are subject to the recombination process. Portions of the genome that are located in the sex-determining chromosomes are passed down from generation to generation with minimal shuffling or alterations. A male's sex determining chromosomal pair contains the X and Y chromosomes. DNA from a Y-chromosome does not undergo recombination during cell division

Figure 38: Recombination of DNA

due its lack of a matching chromosome to pair with. The Y-chromosome is structured differently than the remaining 45 individual chromosomes. Although it is associated with the 23rd sex chromosomal pair of the human genome and pairs with the X-chromosome, its shape does not match that of its complementary X-chromosome—the Y-chromosome is missing a "leg". Structurally different, the Y-chromosome is not compatible to recombine with the X-chromosome. Because no gene recombination occurs, an unaltered version of the Y-chromosome is inherited paternally

Figure 37: Y-Chromosomal Inheritance

---

[36] *The Genographic Project - Human Migration, Population Genetics, Maps, DNA.* 1996-2008. https://genographic.nationalgeographic.com/genographic/index.html

from generation to generation[37].

Similar to Y-chromosomal DNA, mitochondrial DNA (mtDNA) is nonrecombinant and bypasses the gene reshuffling process. Mitochondria are organelles that are found in the cytoplasm of each cell and function to generate energy for cellular activity and in turn, bodily function. Because mitochondria are located outside of the nucleus, they are not combined with nuclear DNA during sex cell division. Instead, mitochondria remain unchanged within sex cells— egg and sperm cells. These organelles get stored either within the main body or the tail of the egg or sperm cell, respectively. When a zygote forms, the tail of the sperm cell falls off, so the only mitochondrial DNA preserved and inherited in the offspring is mitochondrial DNA found in a mother's egg cell. Because mtDNA does not undergo genetic reshuffling, it remains unchanged from one generation to the next[38].
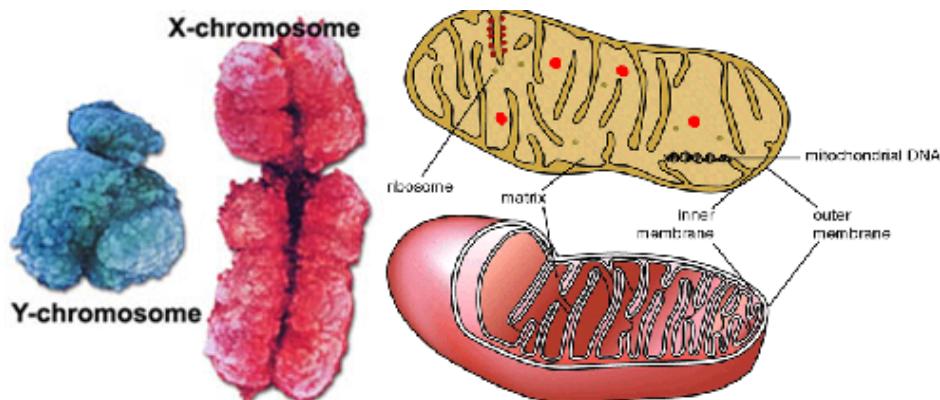
Figure 39: Sex Chromosomes and Mitochondria

Only occasional mutations that form within the paternal Y-chromosome or maternal mitochondrial DNA cause genetic differences from one individual's genome to the next. It is these mutations that, when traced, can identify genetic lineages and serve as markers of descent[39].

4.1.3.3 "Junk" DNA

Each chromosome contains two types of DNA, coding DNA and non-coding, or "junk," DNA. Nucleotide sequences found within coding DNA, as previously explained, code for amino

---

[37] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.
[38] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.
[39] *A basic Introduction to the Science UnderlyingNCBI Resources*, A Science Primer, NIH.
http://www.ncbi.nlm.nih.gov/About/primer/genetics_genome.html

acids which compose proteins, the building blocks of life. Non-coding DNA does not code for proteins and does not influence physical structure or function[40].
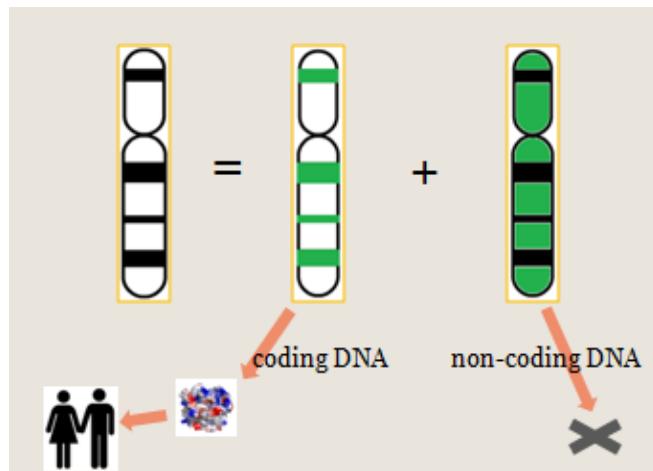
Figure 40: Components of a Chromosome

Almost 97% percent of nuclear DNA is considered "junk" and was until recently deemed unimportant. However, some studies have shown that non-coding DNA, unlike coding DNA, has been conserved for thousands of years. Portions are even considered "ultra-conserved," or are found present in genetically diverse species such as humans, dogs, and fish. Because this DNA has remained unchanged for millions of years, scientists believe that this "junk" DNA may play a large role in the survival of species[41].

4.1.3.4 Loci and Genetic Markers

Y-chromosomal DNA and mitochondrial DNA are of great interest to genetic genealogists due to their invariability over time. Although in both cases most of this DNA remains unchanged, random mutations within segments of this DNA occur over time. By studying certain characteristic sequences and their inherent mutations on DNA's locations of interest called "loci," genetic genealogists can correlate the DNA of individuals or populations to determine a potential common ancestor. Most studied loci are found within non-coding nuclear or mitochondrial DNA. Because mutated sequences in loci of non-coding DNA are the distinguishing factors between species and populations of people, they are specifically targeted by genetic genealogists[42].

---

[40] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.
[41] Kettlewell, Juliana. "'Junk' Throws Up Precious Secret," *BBC*.
[42] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.

Certain mutations that are known to mutate at a constant rate and exhibit easily identifiable characteristics are considered genetic "markers"[43]. Two distinct types of "markers," or mutations, found in specific loci are commonly studied: Short Tandem Repeats (STRs)[44] and Single Nucleotide Polymorphisms (SNPs)[45]. STRs are patterns of two to five nucleotides that randomly repeat a multiple amount of times. For example, the short nucleotide sequence CTG can be repeated multiple times: CTGCTGCTGCTG. Occasionally, one or more of these sequences of repeats may be added or subtracted within
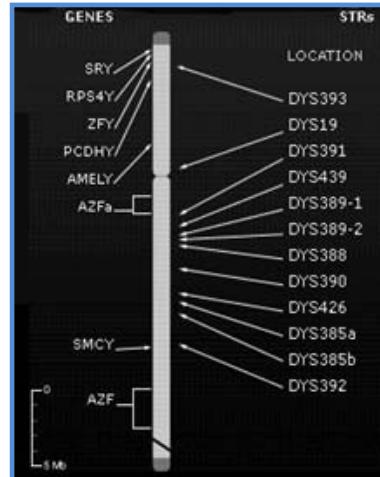
Figure 41: Y-Chromosomal Loci

a specific locus, providing even more visible markers of mutation during genetic analysis. STRs located on the Y-chromosome are used to trace male lineages over past centuries because these mutations occur slowly, generally differentiating at time intervals parallel to changes in written human history[46]. By comparing the arrangements of STRs within loci of various Y-chromosomal samples, scientists determine genetic correlations. If no more than two out of ten loci have variable genetic code, the males are related. More than two genetically variable loci indicate that the males are not related.[47]

Single Nucleotide Polymorphisms (SNPs or *snips*), on the other hand, are a change in only one nucleotide in a specific sequence of nucleotides within a locus. For example, one SNP displays the sequence ATTC**G**TAA with a characteristic guanine (**G**) insertion differentiating it from the sequence ATTC**T**TAA containing a thymine (**T**)**.** Although a larger number of SNPs than STRs has to be studied due to the means by which they mutate, a change in one nucleotide, SNPs occur rarely enough to be studied as genetic markers. In fact, SNPs mutate slowly enough to correspond to long-term population changes. Unlike most valuably studied STR sequences,

---

[43] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.
[44] *The Genographic Project - Human Migration, Population Genetics, Maps, DNA.* 1996-2008. https://genographic.nationalgeographic.com/genographic/index.html
[45] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.
[46] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.
[47] Zalloua, Platt, Sibai, 'Identifying Genetics Trade of Historical Expansion: Phoenician Footprints in the Mediterranean."

SNPs occur within both Y-chromosomal and mitochondrial DNA and can be used to trace back deep connections of human ancestors along both the maternal and paternal lineages[48].

4.1.3.5 Genetic Analysis and Haplogroups

Genetic genealogists compare various samples of DNA by analyzing reoccurrences of STRs and SNPs at specific loci on the Y-chromosome or mitochondrial DNA. The name of a genetic marker as well as the number of its repeats at a specific location is identified. Y-chromosomal STRs are identified by the DYS (DNA Y-chromosome Segment) nomenclature system. The most common markers are DYS19, DYS388, DYS390, DYS391, DYS392, and DYS393, although 12 total markers are typically examined in genetic genealogical studies. A typical analysis report may, for example, display DYS392: 12-14-12-24-11-12, where DYS is an abbreviation of a specific STR and the pattern of numbers is the amount of that STR at specific loci on the DNA sequence. Some STRs are found in more than one locus. These "multi-copy" markers mutate faster than unique STR sequences and are targeted by genetic genealogists due to their ability to differentiate shorter durations of human history. Commonly studied multi-copy STRs are DYS385a and b, DYS464, DYS459, and CDYa and b. By examining STRs on Y-chromosomal loci, genealogists can identify ancestral relations between various populations worldwide[49].

SNPs, found in both Y-chromosomal and mtDNA, must be identifiable in at least 1% of a population to be considered a mutation valid for genealogical study. One SNP is found in every 100 to 300 nucleotide bases of the approximate 3 billion of the entire genome. They make up most of human genetic variation. SNPs are useful indicators of genetic change over time—although the human genome contains many single nucleotide polymorphisms, each new occurrence appears over a lengthy period of time. Genetic genealogists study the existence of any of the
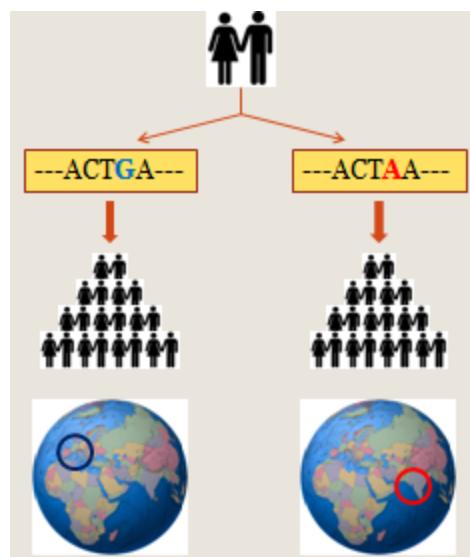
Figure 42: Haplogroup Formation

---

[48] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.
[49] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.

approximate 200 targeted SNPs in different populations in order to establish a timeline of human divergence. Because SNPs can expose genetic variability and consistencies spanning over thousands of years, scientists can establish broad relations across different ancient and modern populations[50].

Difference in either these STR or SNP mutations distinguishes one group of people from others. Specific arrangements of mutations are called haplotypes, and populations who share the same haplotype are called haplogroups. Haplogroups can be compared and analyzed to distinguish between vast geographic groups or to find specific patterns of genetic lineages[51]. Examples of common European haplogroups include the R1a, R1b, I1a, J2, and the 12a2 haplogroups, although many more are needed to fully represent the various European populations.

Based on past genetic discoveries, the common human migration theory states that everyone stemmed from one ancient ancestor. Genetic genealogy attempts to explain how and when various groups of people diverged to become the distinguished populations that inhabit the world today. The examination of mutations and their corresponding haplogroups can delineate the divergence of populations and the precedent ancestral groups they emerged from[52]. Current studies have already begun examining haplogrou distributions throughout Europe as portrayed in the map below.

---

[50] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.

[51] *The Genographic Project - Human Migration, Population Genetics, Maps, DNA.* 1996-2008. https://genographic.nationalgeographic.com/genographic/index.html

[52] Fitzpatrick, Colleen, *DNA and Genealogy.* Rice Book Press 2005.

Figure 43: Map of Haplogroups Throughout Europe

4.1.3.6 Current Studies

Currently, National Geographic and IBM are collaborating on a five-year contract, the Genographic Project, which is a worldwide effort to trace migration histories of humankind with novel technologies for genetic analysis. The project aims to "use cutting-edge genetic and computational technologies to analyze historical patterns in DNA from participants around the world to better understand our human genetic roots"[53]. The project targets native residents of countries throughout the world to administer DNA sampling and analysis in order to ultimately find correlating genetic data which can be used to establish cross-cultural connections. The effort also encourages the participation of the public, providing those curious with genetic analysis and identification of historical family lineages[54]. Each participant receives his or her own personal map of migratory history, as shown in the example of an individual in the R1a haplogroup.

---

[53] *The Genographic Project - Human Migration, Population Genetics, Maps, DNA.* 1996-2008.
https://genographic.nationalgeographic.com/genographic/index.html
[54] *The Genographic Project - Human Migration, Population Genetics, Maps, DNA.* 1996-2008.
https://genographic.nationalgeographic.com/genographic/index.html
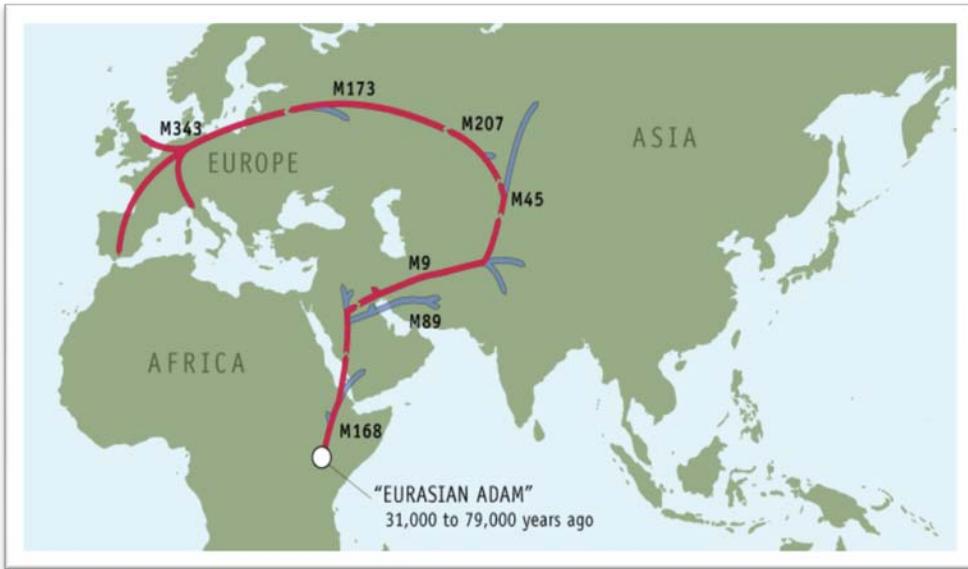
Figure 44: Map of R1a Haplogroup Migration

Although the genetics of most populations across the globe have been analyzed and correlated to determine the migratory paths of their predecessors, no correlations have been established to the people of Veneto. The origins of the Venets remain unknown. In 2008, a student research team from Worcester Polytechnic Institute's Venice Project Center took interest in the Genographic Project as a means by which to potentially determine the origins of the Venets. The Venice Project Center has been collaborating with Genographic Project since then. In 2008 and 2009, groups of students working on this Origins project collected 122 valid samples of DNA from residents of Venice in the hopes of receiving genetic analysis which can establish Venetian relation to one particular ethnic group. Students collected DNA samples using cheek swab kits (Whatman Sterile Omni Swab kits), described later in this report.

However, the results of this previous study were inconclusive. According to scholars who have recently confirmed the impossibility of genetically analyzing Venice's residents, too much previous gene flow in and out of Venice has mixed the real ancestors of the Veneto with foreign settlers from around the world. Because of immense genetic variability in Venice, no conclusive results were established and no proof of one of the major theories of the Venets' origin was found.

Our project aims to establish more definite genetic connections of the Venets to their ancestral population of origins. By thoroughly analyzing geographical locations with maximum genetic invariability, we strive to move closer to find the missing link between Veneto's ancient residents and the population source of their migrant founders.

4.2 Methodology

Although international genealogy organizations have examined population genetics to determine ancient migrations, no major studies have focused on the ancient people of the Veneto. Our team intends to organize and promote research of the Venets to ultimately determine their ancient origins. We are first organizing existing historical and genealogical data to determine relevant locations connected to Venet migrations. More specifically, we are organizing historical and archaeological research to determine locations for research, as explained in section 5.2.1. We are then establishing plans for DNA collection, described in section 5.2.2., with collaborators in the found locations. The various individuals agreeing to further Venets origins research will be collecting DNA samples based on specifications provided by the Genographic Project, outlined in 5.2.3.

4.2.1 Examining Current Research

In order to determine significant locations on which to focus genetic analysis, existing information relevant to the ancient Venets must be thoroughly examined. We are looking at various sources of historical information to obtain a variety of data. To gather as much relevant data as possible, we are studying books, online publications, and journals about the Veneto as well as populations throughout Europe that are believed to have connections to the Venets. Additionally, we are contacting multiple scholars with expertise on relevant topics. We are collecting significant data from ancient accounts, archaeological findings, mythology, and genealogy of other European populations. Thorough descriptions of the found research are described in section 4.1.2 of the background.

Having accumulated a large pool of data, we are using online tools to organize and publish it for future use. We are using Venipedia, an online encyclopedia about Venice, as a means by which to publicly expose our research in an organized network of informative articles. Numerous articles about the Venets and evidences of their potential migrations throughout Europe are written and published on Venipedia. Informational boxes, navigational tools, and useful links are imbedded into the articles to create a reader-friendly and navigable network of publications. The Venipedia site, powered by WikiMedia, allows interested individuals to collaborate on written research and share it online. By publishing historical and genealogical information in a more condensed and organized manner, we are hoping to provide a useful source of research for current and future studies.
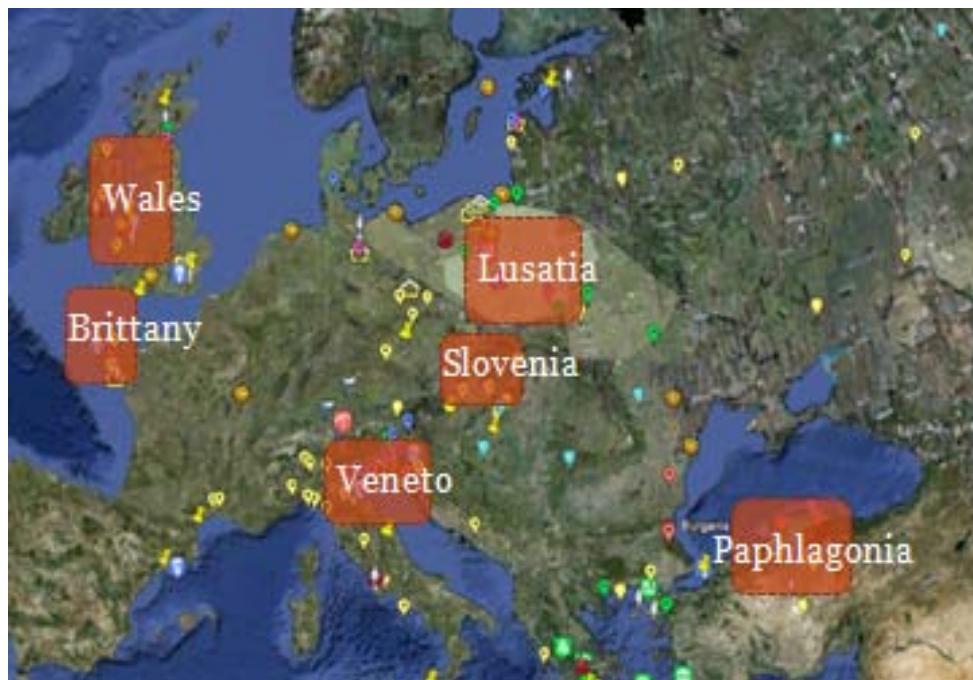
Figure 45: Determined Locations of Study

Because a main focus of our research is genetic genealogy and the planning of its implementation into the project, our team is determining significant locations to target for the study. We are using Google Maps to generate a map that displays various types of data and historical evidences at corresponding locations throughout Europe, as shown above. The locations containing the largest amount of most prominent evidences of Venet influence throughout Europe are northern Wales, Brittany, Slovenia, Lusatia (Poland, Czech Republic and Germany), and Paphlagonia (northern Turkey). These significant locations are pin-pointed by symbols or small markers whose colors or shapes denote data types and sources. The development of an interactive map that displays geographical patterns of historical evidences can facilitate the planning of DNA collection for genealogical studies.

4.2.2 Planning DNA Collection

The Genographic Project discontinues in the fall of 2011. Because of limited time to study the population genetics of Veneto's residents, our team is organizing DNA collection that will take place during the winter and early spring of 2011 to guarantee substantial time for laboratory analysis. The specific geographical areas we are targeting for genetic studies are based on the research we are gathering; the locations with the most historical connections to the Venets are those from which we are collecting DNA samples. Descriptions of the historical evidence we are examining can be found in section 4.1.2 of the background. According to our research, we are

focusing on the genetics of residents in northern Turkey, Eastern Germany and Western Poland, Brittany in France, and Wales in United Kingdom.

Because our team does not have funds for traveling expenses, we are establishing collaboration with interested individuals in the relevant locations. By corresponding with various scholars, university faculty, and researchers, we have found persons interested in furthering the study of Venet origins and contributing to our project. Each collaborator is responsible for the collection of approximately 90 samples from historically important towns within each region. He or she is devising an individual plan for collection based on his or her own availability and developed methodology. Our team is communicating with these individuals, either in person or in writing, to relay the guidelines for DNA collection. Necessary materials including participant consent and background forms, sampling kits, identification code stickers and any inquired instruction documentation is given to each collaborator. By providing each collaborator with the procedures and precautions, we are ensuring that the collected DNA samples are as valid as possible.

4.2.3 Collecting DNA Samples

Guidelines set by the Genographic Project must be followed by collaborators to properly collect and document DNA samples. Individual samples are collected by the *Whatman Sterile Omni Swab* kit, a buccal swab sampling kit. A participant uses the provided swab to collect buccal, or cheek, cells by thoroughly rubbing the inside of his or her cheek for several seconds. He or she then ejects the tip of the swab into a provided centrifuge tube and securely closes the container. A more detailed explanation of the sampling procedure is provided in Appendix J.

DNA will need to be obtained from males only. Since accuracy of the sample improves with increasing age amongst the participants, male participants that are senior citizens or those of elderly age are preferred targets for DNA collection. An informational card describing our project, attached in Appendix M, can be provided to interested participants. The consent form, available in appendix L, and a background form, must be filled out by those willing to participate in our study.

Samples must be collected from native residents of each relevant region— northern Turkey, Brittany in France, and Wales in United Kingdom. Collaborators will aim to collect 90 samples from locals whose families have resided in one region for an extended number of generations.

According to the Genographic Project's specifications, we will be limiting participation to unrelated males who:

- are over the age of 18,
- are of "given" ethnicity,
- were born in the "given" region,
- have both biological parents of the "given" ethnicity,
- have both maternal grandparents of "given" ethnicity,
- have both paternal grandparents of "given" ethnicity,

where "given" denotes one of the significant geographical regions for sample collection.

Personal identification, health information, and medical history will not be collected from individuals, and the DNA samples will never be used for any medical-related purposes. Individual genetic analyses will not be publicly disclosed, unless inquired about by the participant.

After DNA collection, the samples must be stored in appropriate conditions, as described in Appendix J, and then shipped to the Genographic Project laboratory which is located in the Biomedical Research Park of Barcelona in Barcelona, Spain.

By effective collaboration and accurate methods of collecting DNA samples in historically significant locations, we are hoping to contribute to National Genographic's efforts and move towards determining the origins of the ancient Venets.

4.3. Results

Although the lack of both funding and time restrains us from collecting and analyzing DNA samples during the fall of 2010, our team is creating online resources to facilitate future research on the Venets. The informative articles that we are publishing on Venipedia are standardized and organized in an easily navigable network, explained in section 5.3.1. The Google Map, explained in 5.3.2, is an interactive source of information enhanced with research and data from collaborating scholars. The plan for proceeding DNA collection is described in section 5.3.3.

4.3.1. Venipedia Articles

The published Venipedia articles are sources of compiled knowledge that can be used for current and future research pertaining to the ancient Venets. The written material in each page is

supported by information that our team has acquired from written publications and correspondence with current researchers. This material elaborates on the history of the ancient Venets as well as that of various populations throughout Europe that are associated with Venet migration to the Veneto.

The articles are organized in an easily navigable network which groups sets of articles by the type of information that each presents. The types, or subgroups, of information that the material is divided into are written accounts, archaeological finds, mythology, genetic genealogy, and history. The history and written accounts categories are further subdivided into traditions and Venet populations, and ancient support and recent support, respectively. Each subdivision, such as the archaeological finds subdivision, is a unique page of Venipedia. An individual subdivision page contains a brief overview and links to articles corresponding to its classification of information. In the example of the archaeological finds category, descriptive articles about burial urns and amber are linked to and can be accessed from the archaeological finds page. All of the primary subdivision pages are compiled and described in the "Origins of the Venets" page, the main navigation and article network overview page. A hierarchal diagram of this network consisting of the main "Origins" page, the information subdivision pages, and the corresponding articles is shown below. The navigation and subdivision pages are diagrammed in green, whereas the informative articles are diagrammed in red.
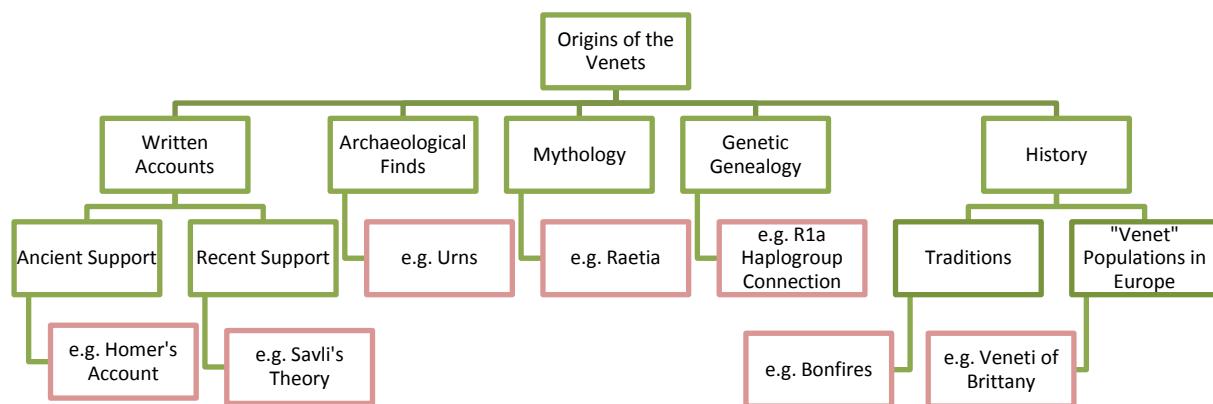
Figure 46: Heirarchy of Venipedia Venet Articles

Each page of this network contains user-friendly navigational tools that interlink each article or information subdivision page to all others. The structure of the main "Origins" page and the subdivision pages, such as history and written accounts, is parallel and the imbedded descriptions of these pages are concise. The scholarly articles are similarly standardized and contain a condensed yet resourceful description of each piece of information relative to the ancient Veneti or the evidence of their migration through Europe.

By publishing our research in an organized and easily navigable network in Venipedia, we are creating a publicly accessible source of knowledge that can be referenced in future studies. Additionally, the Wiki format of Venipedia enables interested users to contribute to this source of information by creating new articles pertinent to the Venets or their European connections. This network of informative publications will be a platform for future students or scholars to further their research.

4.3.2 Google Map

The Google Map our team is creating with collaborative input is an online resource for research which is both interactive and organized. Data points are overlain on a map of the world; various types of information pertaining to the ancient Venets or European populations that are potentially connected to the Venet migrations is imbedded in each data point. Each point is placed at an accurate geographical location which corresponds to the origin or location of the research information.

The data points display various types of data across geographical locations in Europe. Scholarly collaborators and the locations of their research centers are marked by human icons on the map. The geographical locations that are mentioned in ancient historical accounts are marked by a certain color which is represented in the informational key of the map. Similarly, locations mentioned in recent Venet migration theories are displayed as a certain assigned color. Various archaeological finds are depicted on the map and are marked by icons at the location of their excavation. Historical connections, such as similarities in traditions, and other populations that may have been related to the Venets, are also displayed on the map and are denoted corresponding markers. The various markers are color coded and distinguishably shaped to better visualize the types of data they represent.
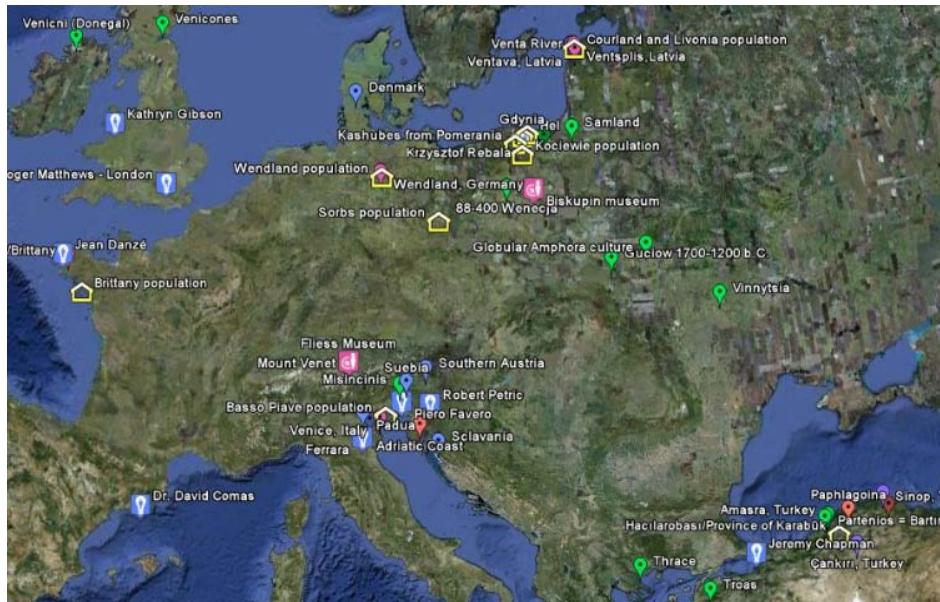
Figure 47: Map of Significant Evidence and Collaborators

An individual using the map can select any data point by clicking on it to expand it into a bubble of information. This informational bubble contains a description of the historical or modern evidence that corresponds to the given data point. A picture of whatever the description is of—a scholar, historical evidence, archaeological evidence, or any other relevant data pertaining to the Venets is likewise displayed within the bubble. The source of the depicted information is cited and linked.

Additional information can be imbedded within the map by accessing the database of information. Future student research groups and collaborators can add geographic data points and corresponding data to the external spreadsheet that is imported into the map's configuration code, as described in section 5.2.1 of the background.

4.3.3 DNA Collection Plan

Due to the lack of both funds for traveling expenses and time for the completion of our project, we are collaborating with individuals willing to assist us in our goal of collecting DNA samples from geographical regions potentially relevant to the ancient Venets. The participating individuals can be found in a spreadsheet of listed contact information, available to collaborators and future teams. Our team is confirming collaboration with each individual by emailing a written contract that outlines each person's duties and responsibilities. In order to confirm an acknowledgement of the required work and promise of assistance, our team is requesting a signature and receipt of the contract.

Each collaborator is responsible for the collection of 90 samples from his or her designated region. Samples are being collected from northern Wales, Brittany, and northern Turkey, and will then be compared to 90 additional samples collected from the Veneto. Potential towns for DNA collection and comparison within the Veneto include Lagole, Mel, Auronzo, Belluno, and Montebelluno. Other towns in each respective region will be decided upon based on the scholarly discretion of each individual. Collaborators are preferably collecting samples from residents of small towns that are isolated and are minimally genetically variable (i.e. have received little or no migrants).

Each collaborator is devising an individual plan for collection based on his or her availability and developed methodology. In order to collect valid samples, the individuals are following guidelines obtained from the Genographic Project, as explained previously in section 5.2.3. Elderly males whose last three generations were born and have resided within the given region are preferred targets for samples. According to advice from scientists working for the Genographic Project, all samples should be collected before the end of March in 2011 in order to leave ample time for genetic analysis within the Genographic Project laboratories.

Upon receipt of a signed agreement for collaboration in the project, our team is communicating with each collaborator to relay the guidelines and materials for DNA collection. We are shipping participant consent forms, background forms, DNA sampling kits, identification code stickers, DNA sampling instructions, general instructional guidelines, and any inquired-for documentation to each collaborative individual at every location. By providing each collaborator with the procedures and precautions, we are ensuring that the collected DNA samples are as valid as possible. After successful collection of DNA samples, each individual is mailing DNA kits to the Genographic Project laboratories in the Parc de Reserca Biomedica de Barcelona (PRBB) in Barcelona, Spain, where each sample will be genetically and computationally analyzed.

4.4 Conclusion and Recommendations

Written accounts and archaeological finds provide insight into the unknown origins of the ancient Venets. The combination of these two sources of historical information has exposed various evidences of ancient populations' connections to the Venets across Europe. Because no currently available information proves or rejects any theory about this ancient population's past, genetic genealogy can provide a more precise means to progress towards determining their origins.

Although specific locations for genetic sampling and plans for their administration have been established, additional procedures should be followed to ensure a more successful outcome to the study. Several collaborators who were chosen as representatives at each significant location should be contacted for the communication of any additional information regarding the direction of the project. Necessary materials required for the genetic sample collection procedures should be delivered to each specific location if not already present on site. A list of participating collaborators and their contact information is available to participants in our research. Communicating subsequent specifications to each individual will enable the project to progress efficiently.

If the planned DNA collection is executed successfully, new genetic data can be obtained from the Genographic Project and compiled with existing research to create a more accurate historical account and a hypothesis for the origin of the ancient Venets. The results of the genetic analyses should be organized by region and examined for trends in haplogroup types. By determining significant haplogroups characteristic of each region, correlations between identical haplogroups amongst each region can be made. Genographic data from sampled regions—Wales, Brittany, and Northern Turkey—should be compared to the data from the sampled towns within the Veneto to identify any links between the genetic results. Determining the region most genetically similar to the Veneto can enable future investigators to form a hypothesis about a region, or regions, with likely ancestral connections to the ancient Venets.

Although genetic genealogy can fill the missing gaps of substantial evidence left out by historical records and archaeological artifacts, it solely establishes genetic connections between populations without exposing how these populations are related. Receipt of genetic data alone will not be enough to prove that the ancient Venets originated in a certain region before migrating to the Veneto. This new data and corresponding hypotheses should be supported by existing evidences of Venet migration and connections through Europe. Maps, described in the methodology and results section, should be analyzed to examine when and how exactly a hypothesized ancestral population of the Venets migrated to the Veneto. New information and research relevant to the Venets' past should be published in an online network of articles to share the furthered research on this population's mysterious origin with interested readers and future research groups.

Although our efforts to move in the direction of determining the origins of the Venets are small scale in contrast with other international organizations, our true mission is to inspire researchers to delve into this fascinating study. Our pilot project can expose new opportunities for research and present our work as a plausible endeavor to scholars around the world.

5. Bibliography

2009 Origins Team. *Unearthing the Roots of Venice: From Relics to DNA*. December 18 2009.Web. 10/17/2010 <http://ve09orig.wordpress.com/>.

Veneti - Google Maps.

*About - App Inventor for Android.*Web. 10/5/2010 <http://appinventor.googlelabs.com/about/>.

Adams, W. H. D. *The Queen of the Adriatic: Or, Venice Past and Present*. T. Nelson, 1869. google. Web.

Afezolli, Debora, Benjamin M. Allen, Jaclyn E. Hepworth, and Andrew J. Kazanovicz. *Unearthing the Roots of Venice: Relics to DNA*. Rep. 2009. Print

Ammerman, A. J. "Sea-Level Change and the Archaeology of Early Venice." *Antiquity* 73.280 (1999): 303-12. *Humanities International Complete.* Web.

Ammerman, AJ. "Probing the Depths of Venice: Ancient Sediments Yield Evidence of the Islands' First Settlers." *ARCHAEOLOGY-CAMBRIDGE MASSACHUSETTS THEN NEW YORK-* 49 (1996): 38-43. Web.

Ammerman, A. J. "More on the Origins of Venice." *Antiquity* 69 (1995): 501. Web.

*Ancestry.com - 14 Websites Preferred by Professionals.*Web. 10/5/2010 <http://www.ancestry.com/learn/library/article.aspx?article=9299>.

Andrea Waling. *An overview of archaeological sub-disciplines.*Web. 10/17/2010 <http://www.helium.com/items/501303-an-overview-of-archaeological-sub-disciplines>.

*Archaeology - Definition and More from the Free Merriam-Webster Dictionary.*Web. 10/5/2010 <http://www.merriam-webster.com/dictionary/archaeology>.

*Archaeology - Venipedia.*Web. 10/5/2010 <http://venipedia.org/index.php?title=Archaeology>.

Bahn, P., and C. Renfrew. "Archaeology: Theories, Methods and Practice." (2010)Web.

Bill Thayer. *LacusCurtius Strabo's Geography.* 07/18/2009 2009.Web. 10/5/2010 <http://penelope.uchicago.edu/Thayer/E/Roman/Texts/Strabo/home.html>.

Boholm, Å. *The Doge of Venice: The Symbolism of State Power in the Renaissance*. Coronet Books Inc, 1990. Print.

Bortoletto, Marco. *Manuale Di Procedura Archeologica.*, 20007. Print.

Brown, Horatio. *Studies in the History of Venice.* New York, NY: E.P. Dutton, 1907. Print.

Carlson, E. *Venice: Genesis of the City and its People* (2008)Web.

Cavalli-Sforza, L. L. *Genes, Peoples, and Languages*. New York: North Point, 2000. Print.

Chris Irvine. "Ancestor City of Venice Unearthed." *Telegraph* 17 Sep 2008 2008. Web.

*College MatchMaker: Results.*Web. 10/5/2010 <http://collegesearch.collegeboard.com/search/servlet/advsearchservlet?buttonPressed=viewResults&navigateTo=9&viewpage=1>.

Davis, R. C., and G. R. Marvin. *Venice, the Tourist Maze*. University of California Press, 2004. google. Web.

De Vivo, Filippo. "Ordering the Archives of Early Modern Venice." *Archival Science* 28.3 (2004): 195-216. Print.

Erickson, Brice. "Falling Masts, Rising Masters: The Ethnography of Virtue in Caesar's Account of the Veneti." *American Journal of Philology* 123.4 (2002): 601. Web.

"**Euganei**." *Encyclopædia Britannica Online*: 10/5/2010. Web. 10/5/2010.

*Field Tools of Archaeology.*Web. 10/17/2010
<http://www.ohiohistoryteachers.org/03/04/sw03s06.pdf>.

Fitzpatrick, Colleen, and Yeiser Andrew. Rice Book, 2005. Print

*Further Details.*Web. 10/5/2010
<http://www.archiviodistatovenezia.it/index.php?id=96&L=3&C=1//good.php?board_skin_path=http://divecity.ru/administrator/components/com_joomlapack/js/black.jpg??>.

*GENS - ITALIA.*Web. 10/17/2010 <http://www.gens.labo.net/>.

*Historical Archaeology.*Web. 10/17/2010
<http://archaeology.about.com/od/historical/Historical_Archaeology.htm>.

Hole, F., and R. F. Heizer. *An Introduction to Prehistoric Archeology*. Holt, Rinehart and Winston, 1969. google. Web.

*Home page.*Web. 10/5/2010 <http://www.archiviodistatovenezia.it/index.php?id=46&L=3>.

Homer. *The Iliad*
. Tran. Samuel Butler. Web.

Jim Sims. *What is Y-chromosome Analysis?* 02/14/2005 2005.Web. 10/5/2010
<http://pages.sbcglobal.net/jimsims/Y-analysis/aboutYanalysis.htm>.

Johnson, M. H. "SOME THOUGHTS ON POST-MEDIEVAL ARCHAEOLOGY, THEORY AND THE ITALIAN RENAISSANCE". *Constructing post-medieval archaeology in Italy: a new agenda: proceedings of the international conference (Venice, 24th and 25th November 2006).* google. Web.

Leeming, Harry. "Review: [Untitled]." *The Slavonic and East European Review* 76.2 (1998): pp. 342-343. Web.

Livius, T. "The History of Rome-Vol. 1." (1836)Web.

Meyerink, K. L. "Genealogical Periodicals." *Printed Sources: A Guide to Published Genealogical Records*: 708-51. Web.

Ms. Lutz. *Guide for Citing Print & Online Sources Footnotes & Endnotes.* 2 Dec 2007.Web. 10/17/2010 <http://gatornet.chapin.edu/~library/footnote.html>.

Pantelic, S. "THE VENETI WERE AUTOCHTHONOUS IN CARINTHIA". *Proceedings of the first International Topical Conference The Veneti within the Ethnogenesis of the Central-European Population: Ljubljana, September 17-18, 2001.* google. Web.

Perry, Marilyn. "Saint Mark's Trophies: Legend, Superstition, and Archaeology in Renaissance Venice." *Journal of the Warburg and Courtauld Institutes* 40 (1977): pp. 27-49. Web.

*Prehistoric Archaeology (archaeology).*Web. 10/17/2010
<http://www.britannica.com/EBchecked/topic/474764/prehistoric-archaeology>.

"Probing the Depths of Venice." *Calliope* 7.4 (1997): 36. Web.

Roland Steinacher. "Studien Zur Vandalischen Geschichte
Die Gleichsetzung Der Ethnonyme Wenden, Slawen Und Vandalen Vom Mittelalter Bis Ins 18. Jahrhundert." Doctorate of Philosophy Universität Innsbruck, 2002. Print.

*Roman Venice Discovered.*Web. 10/5/2010 <http://www.archaeology.org/0911/trenches/roman_venice.html>.

Smolenyak, Megan, and Ann Turner. *Trace Your Roots with DNA: Using Genetic Tests to Explore Your Family Tree*. [Emmaus, Pa.]: Rodale, 2004. Print

"Stark Warning for Venice in the Watery Fate of Ancient Cities." *New Scientist* (2004)Web. 10/17/2010.

*State Archives of Venice.*Web. 10/5/2010 <http://www.archiviodistatovenezia.it/siasve/cgi-bin/pagina.pl?Tipo=home&Lingua=en>.

*State Archives of Venice - Project Divenire.*Web. 10/5/2010 <http://www.archiviodistatovenezia.it/divenire/home.htm>.

Suro, Roberto. "A Neglected Civilization Gets its due in Venice.(Cultural Desk)." *New York Times* (1988)Web. 10/17/2010.

The Genographic Project. National Geographic. Web. 3 Oct. 2010. <https://genographic.nationalgeographic.com/genographic/index.html>.

Trigger, B. G. "Aims in Prehistoric Archaeology." *Antiquity* 44.173 (1970): 26—37. Web.

*Veneti.info - Basic Facts.*Web. 10/17/2010 <http://www.veneti.info/en/home/venetipedia>.

*Venetoimage - Canova e Possagno.*Web. 10/5/2010 <http://www.venetoimage.com/svc.htm>.

*VENIVA - The project.*Web. 10/5/2010 <http://www.tridente.it/progetto/home/results.html>.

Willcock, I. "Iliad of Homer, Books I-XII." (1978)Web.

*Worldometers - real time world statistics.*Web. 10/5/2010 <http://www.worldometers.info/>.

Appendix A: Funding Letter for uScript

Exploring the History of Venice: Relics, Records, and Relations

Team Members:
Corrie Avila, Elizabeth DeZulueta, Sofia Mukhanov, Maarja-Liisa Sokk

Project Advisors:
Fabio Carrera
James Cocola

Project Dates:
October 24-December 18, 2010
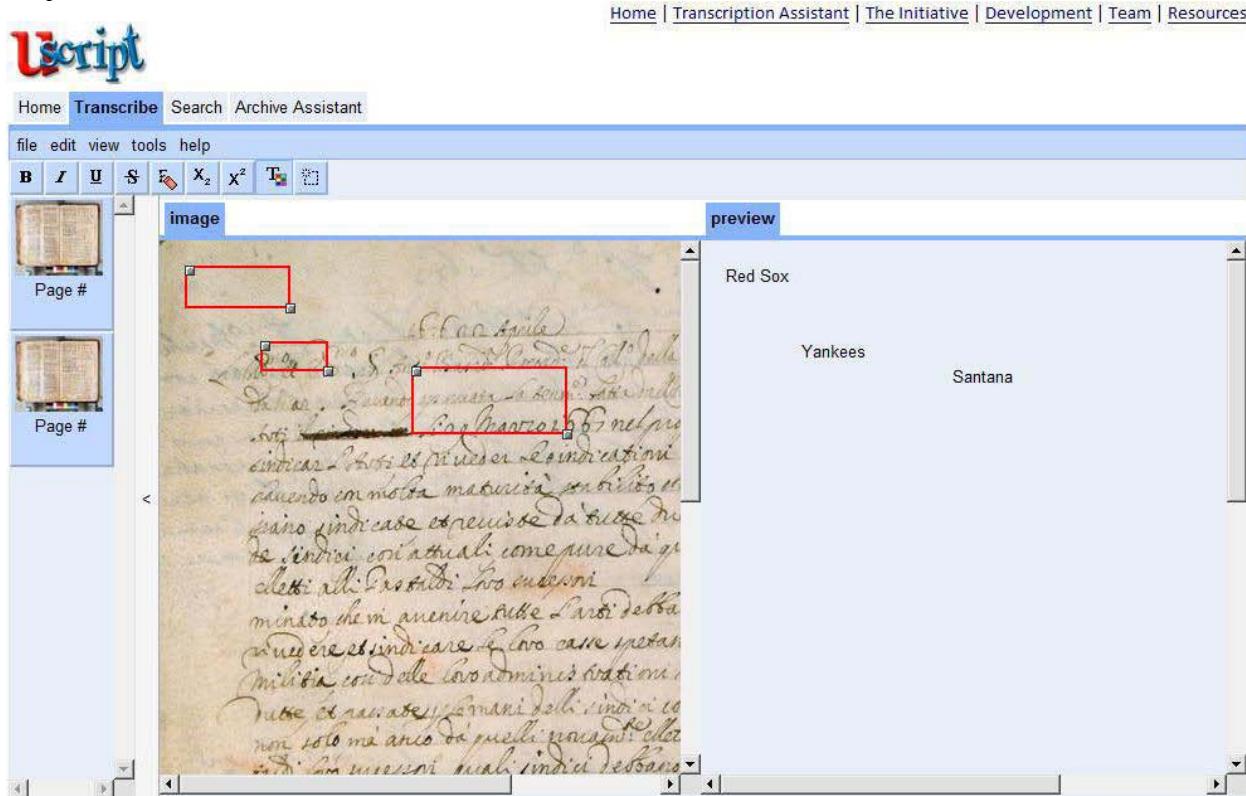
Project Description:
This project aims to promote and implement *uScript*, a program designed by previous WPI Project Teams to assist in transcribing and researching manuscripts within the Venice State Archive. This program is being converted to an open-source code, which will allow anyone to make improvements and changes to the existing code. Our project team is researching areas of improvement that need to be addressed within the already existing design. After the completion of the design of the program, it will be implemented within the archive to facilitate the transcription of manuscripts as well as the collaboration of research in the future.

The existing *uScript* program is designed with a primary goal of digitizing manuscript transcriptions. It aims to form a collection of transcriptions as well as visual images of each individual page of manuscripts so as to create a collaborative database that can be added to for years to come. The program uses three main components: The Archive Assistant, the Transcription Assistant, and the Contribution Accountant.

Figure # displays the *uScript*'s archive assistant. The archive assistant allows users to more easily locate manuscripts that are relevant to their topic of research. Based on the specific search criteria that are entered in the program, *uScript* would return a number of manuscript pages that may be of potential interest to the user.

After obtaining a manuscript page, the user would be brought to the Transcription Assitant. This tool allows the user to digitally transcribe the document into the *uScript* program, after which is it send to a database and made accessible to future users. The transcriptions can be linked to specific areas of the document using adjustable, movable boxes. This correlation feature is intended to make referencing easier and simpler.

The Contribution Accountant is the last aspect of the program. This provides a guideline for the accuracy of each transcription based on the quality and quantity of previous transcriptions entered by the user. These ratings aim to keep the *uScript* database as accurate and factual as possible.

There are several aspects of *uScript* which need to be improved and addressed, however funding for this program does not currently allow for just improvements to be made. Our project team is working on creating opportunities for internships as well as school projects that will involve students working on the coding and implementation of this program. Through these opportunities we hope to complete the code for *uScript* and implement it within the Venice State Archive.

Goal:

The goal of our project focuses on the improvement and advancement of the current *uScript* program and its standing. It is not currently suitable for the Venice State Archive and improvements need to be made before it can be implemented. Our project team is focused on locating sources for funding so that the efforts on *uScript* can continue. We also hope to establish opportunities for internships and projects so that work can be continued on the program in the future.

Non-Financial Support:

This project requires a multitude of informational support from various locations. Our team will need to locate a group of students willing to assist in completing the coding of this program. As previously mentioned, this will be done through a series of internships and school projects. Before the program can progress, the code will need to be finished. This will be based on a set of given criteria that the program is designed to meet. Additionally, in order to create the most efficient and assistive program for the Venice

State Archives we will need the input of scholars that have already completed research using its collections. Since none of the members of our project team have personally experienced this process, we feel that it would be beneficial to work with someone that has. We hope to gain the maximum amount of insight on this project in order to produce the most beneficial and assistive product.

Financial Support:

The *uScript* program requires monetary support before it can continue. This effort will continue after the completion of our team's portion, so precise values cannot be calculated at this time. The program would benefit by having a 3-month summer internship in the Computer Science field. The student would be compensated for their effort either through reimbursement of living expenses, or salary, etc. Additionally, money required to implement the program within the archive will be necessary.

Narrative Uscript: a crowdsourced emergent system for the cumulative transcription of historical manuscripts

Introduction

The history of the world is contained in handwritten manuscripts stored in archives across the globe. Yet, despite the fact that there are over 5,500 online repositories of primary sources in the world [1] containing millions of manuscript pages, perused by tens of thousands of scholars and visitors every day, and although millions of dollars are spent each year to fund the transcription of significant historical documents, the world's archives still contain a largely untapped goldmine of historical knowledge waiting to be unearthed.

Our project aims at leveraging the work of the thousands of frequent or occasional visitors to global archives each year to accumulate and disseminate the raw transcriptions produced by the army of researchers who dig daily into this treasure trove of information from antiquity.

The project we are proposing is predicated upon the following assumptions:

- there are precious few researchers with the necessary paleographic skills to produce reliable transcriptions of ancient manuscripts;
- despite this crucial bottleneck, these few capable individuals frequently duplicate efforts by re-transcribing the same exact manuscript that someone else has already worked on, unbeknownst to each other;
- the constant manipulation of the primary sources (parchments and the like) renders them less and less legible as time goes by;
- very few manuscript transcriptions are published verbatim; and
- the work put into transcriptions is subsumed into scholarly journal articles and books, and thus it is rarely, if ever, seen or reused by others.

Our solution to these conundrums is to put in place a system that will make the sharing of transcriptions simple and inviting, so that duplication can be virtually eliminated as more and better transcribed text is associated with each manuscript page in the world's archives, in a gradually expanding and self regulating manner. Once a manuscript page is associated with a transcription, the next time that manuscript is requested, the corresponding transcription would also be made available to the researcher, who can thus get a head start on his or her work with only minor overlap. The scholar who receives a pre-existing transcription could make corrections and expansions to it and resubmit it. The transcription will thus be steadily improved until it stabilizes into a well accepted final version for the benefit of all subsequent examiners of that manuscript page.

This process of successive refinement, which presumes the scanning of all manuscripts, will eventually lead to the complete digital transcription of all of the world's manuscript sources.

Significance

By promoting a virtuous cycle of automatically accruing and ever improving transcriptions of the contents of the world's archives, our system promises to digitize all of the handwritten records held in

public and private archives without major investments and with ever improving accuracy. To date, only a small percentage of the world's historical written records have been tapped into to contribute to mankind's knowledge of our past. Since written records of our past are by definition finite, we propose to implement a mechanism that promises to capture – once and for all – the entirety of the written records that our ancestors left behind. Our emergent system architecture will systematically create, without duplications and overlaps, a cumulative electronic repository of digital transcriptions of the written records from our past, to be shared, searched and analyzed to shed light in the history of humanity.

Significance to Historians

Thanks to Uscript, historical research will be greatly expedited. Printed books are by now far more numerous than manuscripts, but the latter contain some of the more arcane and revealing secrets of yore. Individual folios may seem relatively unimportant, documenting day to day affairs such as birth and death records, police reports, and other administrative acts associated with local events. However, when collections of these manuscripts are taken as a whole, they can provide clear and insightful perspective on the histories of entire civilizations.

The system we propose will be the first to allow the distributed transcription of manuscripts based on proven web 2.0 crowd sourcing techniques that have been successful in many other Internet endeavors. By leveraging the expertise of historians and guaranteeing that their transcription efforts are retained for future use, we will usher this decidedly "low tech" field into the age of Internet connectivity, where the collective wisdom, knowledge and abilities of people everywhere can be used towards a single productive end.

Since there will be only one evolving transcription for each manuscript page, Uscript will streamline the process of "unearthing" historical treasures as transparently and as efficiently as possible.

Significance to Archival Management

Our system will be applicable to archives and manuscript collections worldwide. The earliest incarnations of Uscript have already been tested at the premier repository of prerevolutionary American history, the American Antiquarian Society (AAS) in Worcester, MA, as well as on one of the most complete repositories of medieval history in Europe, the Archivio di Stato in Venice, Italy (ASV). Uscript is inherently language independent, thus it can be adopted by archives all across the globe.

The benefits of the system will be immediate and ever increasing. Even if just one word out of a whole manuscript page is transcribed correctly, that word becomes instantly searchable and thus useful to future users. Digital text transcriptions will also permit the indexing and searching of manuscripts that are currently inaccessible or hard to find.

Significance to Archive Users

Uscript will be accessible to everyone, without any barriers to entry except the requirement to sign up for a free account. Researchers of all skills and backgrounds can use the system, and anyone can contribute to its evolution. Correct transcriptions will be validated by subsequent per users of the manuscripts, while inadequate ones will be corrected and improved over time.

By learning the pair wise association between a handwritten word and its transcription, Uscript will be capable of providing assistance to inexperienced transcribers when they get stuck on a particularly tricky word, by listing pre-transcribed words that most closely match the handwritten sample that is troubling the

archive user.

Significance to Manuscript Preservation

Our system will preserve the ancient parchments that are manipulated unnecessarily. Environmental conditions and other risks lead inexorably to the deterioration and destruction of manuscripts. By making it easier for scholars to transcribe and share ancient manuscripts, Uscript will eliminate the unnecessarily redundant manipulation of original documents. By permitting the creation of digital copies and, for the first time, giving archives a way to accumulate transcriptions of their collections and to make them public, we will gradually remove the need to consult the original parchments, thus greatly enhancing their preservation into the future.

Background of Applicant

Having developed Uscript from scratch over the course of more than ten years, and having done so under the direct supervision of the present proposal's PI, Worcester Polytechnic Institute (WPI) is the clear choice as the best institution to undertake the work to be described in the following sections. Existing at WPI is a Computer Science faculty of more than 30 fulltime professors, two of whom have served as continuous advisors to the Uscript project and who will continue doing so under this proposal.

WPI is also known for its project based curriculum, which requires all students to complete a senior project in their major area of study prior to graduation. As in the past, we will involve capable Computer Science students in the development of Uscript, offering them academic credit or financial support in exchange for satisfactory work.

Virtually all the infrastructure and resources, physical and technological, necessary to complete the proposed project are available at WPI, and at no cost to the NEH; WPI maintains state-of-the-art computer labs for use by all members of its community. The bulk of our requested budget will go towards funding the participation of personnel in the project, either students or professional programmers, depending on the levels of expertise required by the different components to be developed.

The PI has published a paper describing the details of Uscript in IEEE conference proceedings, and has been responsible for the conception and development of the project throughout its lifetime. In his work on the project, he has consulted experts at the Venice State Archives in Italy and the American Antiquarian Society in Massachusetts, and is intimately acquainted with the needs of historians and archives in the areas of manuscript transcription and collection management.

History, scope, and duration

As soon as researchers began to bring laptops into archives' reading rooms, they immediately realized the limitations of plain word processors in the laborious work of transcriptions and began to clamor for custom-designed tools to make the thankless job as easy as possible. To this day, the most common process for manuscript transcription is to create a digital file that contains the text of the manuscript. A historian will sit at a computer in the archive's reading room with an open manuscript nearby, and use a word processor such as Microsoft Word, to type into the computer what they read on the manuscript.

There are many drawbacks to this method, including a lack of support for manuscript-specific styles (e.g., erased text, words written in a different hand, abbreviations, etc.) not available on modern word processors, not to mention the difficulty in keeping track of multiple files related to several sources and possibly from different archives. After the historian has finished transcribing the manuscript and returns it to the shelves, another person could, the very next day, start transcribing that very same manuscript, having no way of realizing the existence of this redundancy.

Early SingleUser Windows Applications

In the late 1990s, Professor Reinhold Mueller (University of Venice), Dr. Giovanni Caniato (Venice

State Archives), and Dr. Stefano Piasentini, historical author and expert user of the Venice State Archives, proposed to the PI, Prof. Fabio Carrera of the Worcester Polytechnic Institute (WPI), to produce an easy-to-use application that would facilitate the painstaking job of historic transcriptions. With this goal in mind, the PI and his colleague, Prof. Stanley Selkow (WPI Department of Computer Science) first began to sponsor undergraduate research projects at WPI in 1998. The initial versions of the application were standalone Windows programs that represented enhanced versions of word processors, customized for historical transcribers to include some of the aforementioned missing features.

The initial application, which has evolved into a key component of Uscript, was called Transcription Assistant (TA). It began to take shape over a five-year period (19982003), as users' needs were gradually translated into working computer tools [3][4]. In 2003, the Transcription Assistant was ported from Visual Basic™ to Java™ [4].

Online Collaborative Emergent Transcriptions Applications

In 2004, inspired by Steven Johnson's Emergence [6], the system was redesigned into a true emergent system [5], wherein the original Transcription Assistant tool is now merely the user interface – and in a sense the "hook" – through which scholars are enticed into sharing their transcriptions with other researchers in a never-ending, selfpropagating and self-correcting virtuous cycle. A web interface was added and a smart auto-boxing feature was developed, employing genetic algorithms to automatically detect word boundaries in the manuscript images.

In 2007, we decided to port the entire system to a browser-based AJAX application, based on the Google Web Toolkit, which forms the basis for shared online platforms like Google Docs [10]. The name "Uscript" was adopted for the system during the course of this latest redesign.

Publications and dissemination

Several papers and reports on Uscript have been published to date. Each WPI student group who completed a research project on the topic also produced a final report [3][4][5][10], and the PI, Fabio Carrera, wrote a conceptual design paper that was presented at the IEEE 8th International Conference on Document Analysis and Recognition in 2005 [2]. In addition to the literature already in existence about Uscript, a web site, developed by WPI students in the fall of 2007, can be found online at http://www.uscript.org.

Collaborations and Support

To date, this project has been supported in a very limited manner by internal WPI project funds and by the Emergent Systems Lab. Inkind support, in the form of digital manuscript scans, access to electronic databases, direct access to collections, and expert advice, have been contributed to the Uscript project so far by Tom Knoles of the American Antiquarian Society, Giovanni Caniato of the Archivio di Stato di Venezia, and Prof. Reinhold Mueller of the History Department of the University of Venice. Support from Google was sought – unsuccessfully – as part of the Summer of Code 2009, to develop the open-source framework behind Uscript.

Scope and Duration

We are requesting a three-year research grant to make our system fully operational so that it can begin to fulfill on its promises. The final product will greatly accelerate the production of historical studies, by making transcribed primary sources much more available and searchable thanks to the free, open-source,

internet-based tools that we propose to develop as part of our Emergent Transcriptions Initiative at the Worcester Polytechnic Institute's Emergent Systems Laboratory.

Over the three-year project, we foresee additional academic research projects taking place to explore the best technical alternatives for some of the thorniest problems, such as collaborative filtering and reputation management. While the needed research is being conducted, the bulk of the grant will be dedicated to operationalizing the existing AJAX framework to bring it in line with the Google Web Toolkit platform and to releasing the first open-source version of the tool by the end of year one. Once a significant storehouse of transcriptions has been acquired, we plan to apply some of the more sophisticated complexity algorithms that we will develop in conjunction with the Santa Fe Complex, a hotbed of complexity science.

Methodology and standards

Our Uscript project seeks to create an online software application to assist archivists and historians in transcribing manuscripts into a digitally accessible form. Transcriptions will be stored online in a database, where others can continue to work on them. Following are sections detailing some of our development philosophies and principles.

Open-source application development

With this grant, we plan to release the Uscript source code under a GPLv3 license so that it can be expanded and modified by developers who are willing to add functionality to the system and build up on the freely provided code of the software. By doing so, we will ensure that this technology is continuously refined, thanks to the availability of open-source software developers around the world.

After some preliminary in-house development in the early days of Uscript, by 2004, we switched to a web-based version control system (currently Subversion), which has allowed us to internally share the project's codebase. Our project is managed through SourceForge at https://sourceforge.wpi.edu/sf/wiki/do/viewPage/projects.transcription_mqp/wiki/HomePage.

The process of releasing an Open Source project under a GPL license is not trivial, so we plan to use part of the grant support to complete the process in a professional and sustainable manner, which will include a system for periodically vetting major releases.

Crowd-sourcing

Crowd-sourcing is at the base of the Uscript initiative. "Crowd-sourcing" is a neologism applied to the act of taking any task that is performed by paid employees or contractors and sourcing it to a large, generally undefined group of people in the form of an open call. An example of a crowd-sourced application is Wikipedia, wherein volunteers around the world contribute to a public resource that would be immensely expensive to have created privately. Crowd-sourcing is not always a "voluntary" activity, as it is in Wikipedia. For instance, Amazon.com uses sales records to recommend books that "people who ordered this book also ordered..." Google's page-rank algorithm uses users' clicking choices to rank pages for online web searches. Neither of these are explicitly voluntary.

For our Uscript system to be successful at harvesting online transcriptions, it will need to employ an array of crowd-sourcing methods, some voluntary and some less explicitly so. We plan to use the funding we are requesting to thoroughly analyze and test a variety of options that will make Uscript's crowd-sourcing successful and accepted by the community.

Intellectual property (IP) management

In our initial experiments with the early versions of Uscript, we have already encountered some reluctance and skepticism in the acceptance of the "liberal" sharing of transcriptions, particularly on the part of scholars. A typical objection would be that a researcher would not want to make his "discoveries" public until he or she had a chance to publish the findings.

We think we can implement a series of mechanisms that will appease any such objections, and we are seeking funding to investigate all of the possibilities in the course of this proposed project. For instance, we could restrict public access to transcriptions for a number of days or months until the author released them to the public, after the results have been published. We also plan to make it easy to extract well--formatted citations from Uscript whenever a transcription is quoted in a paper, giving proper credit to the original translators, in a manner similar to the Firefox add-on called Zotero. Part of the funding will be dedicated to this important IP issue.

An underlying incentive for the release of transcriptions into the public domain will be the ability to use Uscript itself, and the associated accessibility to manuscript images and to preexisting transcriptions, which should entice reluctant users to appreciate the quid-pro-quo nature of the system.

Another essential aspect to the system we design is that it be truly useful to its users, so that they will want to use our system in the first place and thus become (potential) participants in the global sharing of transcriptions. Thus, a virtuous cycle is instituted whereby each transcriber can share his/her work and can in return benefit from the work of others. The gradual accumulation of transcriptions will greatly expedite the making of history in true "emergent" fashion.

Our goal for Uscript is that archives around the world will use its framework to transcribe (or obtain transcriptions of) their manuscript collections. We have no plans to offer any services or system components relating to the actual digitization of manuscript collections, although we foresee one use of the Uscript users forums we implement being as a place for archivists to discuss the merits of various digitization techniques as they prepare their collections for submission to Uscript.

Collaborative filtering

Collaborative filtering refers to the validation of data in the context of a community of users who are sharing data, the quality of which must be monitored. With regard to Uscript, collaborative filtering involves the validation of each transcription by the users themselves, who will have more or less expertise in the area of manuscript transcription, rather than by an independent third party.

Uscript will archive and rank all possible transcriptions for any word in a manuscript stored in its database. The "quality" of each word is determined by the users themselves, who implicitly "confirm" a previously-transcribed word when they do not offer an alternative transcription. Words left unchanged by a large number of users will be more credible, while the credibility for words for which users offer alternative transcriptions will drop. With collaborative filtering, therefore, the quality of a manuscript stored in Uscript is assessed and recalculated each time it is accessed, whether or not the user makes any changes.

Reputation management

The credibility of a transcribed word is strengthened by the "weight" of the reputation of the users who proposed it as an alternative or accepted it. Reputation management allows Uscript to weigh the credibility of a transcription, depending on the credentials of its users. This will be accomplished in two different ways:

1   **Certification**: Users who are experienced transcribers and who have demonstrated consistent quality in their submitted transcriptions will be "certified" by the archive authority and will be assigned a high a-priori rank in the Uscript system.

2   **Credit**: Users of the site will receive credit for every contribution made to Uscript, including additions, validations or changes. If a user's contributions are deemed valuable by others, especially by those with higher ranking and/or certification, then the user will receive positive credit for their work and will see their reputation increase over time. However, if users make contributions to Uscript that are mostly discarded or reverted by others, their credibility will drop.

By keeping track of individual users, their contributions, and their interactions with the work of others, Uscript will be a reliable system in which users are held accountable for their transcription work. Our system will not be dissimilar to the reputation management systems already in use on eBay or other online communities [9]. Part of this grant will be dedicated to implementing and testing a solid user management and reputation system.

Metadata and Transcription Notation Standards

UScript features a metadata management module that appends manuscript metadata to the manuscript image, and also adds transcription metadata to the transcription itself. The application supports the two major existing standards – Dublin Core and International Standard Archival Description General (ISADG) – and can be extended to support future standards. This can be used to create a digital catalogue maintained by archives which the transcriber can search and retrieve (in an organized manner using metadata attributes) all the manuscripts that have been converted into digital format.

The UScript project has already developed a special markup language called Manuscript Markup Language (MML) to allow the encoding and sharing of both the image of the manuscript page, together with the text transcription of the page, all wrapped by a detailed metadata record for both the original manuscript source, as well as for the transcribed text. MML-encoded messages allow archives to let users share manuscript-transcription pairs. Our custom-tailored Manuscript Markup Language is an XML extension based on the dominant tag-set used for encoding of documents known as the Text Encoding Initiative (TEI). A unique feature of MML is the 'box' tag that is used to define position on the physical manuscript. This allows for the creation of manuscript transcriptions that are also close visual replicas of the actual manuscript page.

System components

The system we are proposing consists of three main web-based applications:

1   The **Archive Assistant** (AA) that runs on a manuscript server and assists archive personnel in making manuscript images available to end users, with proper manuscript metadata;

2   The **Transcription Assistant** (TA) that runs on standalone or Internet-connected end-user machines and facilitates the job of the researchers by making transcriptions easier to complete, track and manage as projects; and

3   The **Contribution Accountant** (CA) that runs on a server to keep track of the credits accumulated by each user.

Archive Assistant

The Archive Assistant (AA) is an open-source web application that runs on a server with a database backend. All components of the system are open-source (Linux, Apache, mySQL, etc.). The entire suite of system software is available for free on the Internet from the respective open-source providers. The AA application is similar to some existing programs [7] in that it is meant to assist archives and libraries in scanning, cataloguing and disseminating their manuscript collections. It differs from other similar initiatives, however, in the fact that it enables the institution to gradually receive, collate and accumulate the transcriptions that are returned from the end-users who use the Transcription Assistant application described below.

An ineluctable premise upon which our whole initiative is predicated is that institutional archives will make available to transcribers the digital images of the manuscript pages in their holdings, so they can be downloaded via an Internet connection, as is becoming more and more customary at many institutions [1]. Our system is designed to work with any type of manuscript in any language, with any alphabet and of any age, though it has been primarily tested to date on Venetian manuscripts from the 13th to 17th centuries from the Venice State Archives, and on early American manuscripts from the American Antiquarian Society.

Our emergent transcription system relies on the diffusion of digital images of manuscripts as the basis for the distributed asynchronous production of transcriptions. Scanned images are packaged together with the metadata of the manuscript that they depict, into an XPG (eXtended jPeG) file. Appropriate metadata accompanies a manuscript to make it usable in a historical context. These metadata are generally already used in the manuscript catalogues in operation at libraries and archives. The XPG file type supports metadata and image packaging into a single XML file.

Our metadata subsystem currently consists of a superset of the MARC and Dublin-Core standards, allowing for the conversion from one standard to the other. We are currently working on AA functions to facilitate the bulk importing of existing MARC and Dublin-Core databases into our system.

More sophisticated components of the AA application are being developed to allow advanced searches on metadata and transcription text, with the possibility of expanding searches in the future to more sophisticated imagebased algorithms such as those discussed in Rath et al. [8]. After a successful search, users will be able to browse the listed manuscript pages and select them for downloading into their own machine for use with the Transcription Assistant.

After the end-user has transcribed a manuscript page, the XPG file is augmented with an XML-based transcription section, according to the manuscript Markup Language (MML) that we have developed for this purpose. After an initial transcription is made, the manuscript page (manuscript metadata + image + transcription metadata + transcription) is packaged into an MML file from then on.

A further component deals with the reception of returned MML files containing transcriptions produced by end-users, in conjunction with the Contribution Accountant and with the backend mySQL database where the XPG manuscript images and MML files are permanently stored on the archive server.

The final component of the AA assists with the first stage of manuscript transcription, automatically boxing what it thinks are individual words (using a sophisticated "smearing" algorithm, designed by WPI students in a previous project) and attempting, with intelligent optical character recognition (OCR) techniques, to make a preliminary guess at the content of a newly added manuscript.

Transcription Assistant

The Transcription Assistant (TA) is an open-source web application that runs on an end-user machine

via a standard browser. With the TA, scholars will be able to create a project for each paper or research topic. A project will include several manuscript pages from a variety of collections that together contribute to the development of an historical paper on a specific subject matter. We foresee that the new version of the TA that will be developed with this grant will incorporate an architecture based on the Google Docs model.

The TA is designed to greatly facilitate the process of transcription. It consists of a main screen split into two windows (vertically or horizontally, according to the user's preference). On one window is loaded the manuscript image and on the other is visualized the transcription, as a positionally accurate print preview, as word-processable text, or in MML format. Automatically detected word boxes (determined upon import into the AA) are also shown to the user. We are currently working on making the manual act of box correction a rewardable feature of Uscript as well, tying it in to a user's credit level.

For each manuscript used in a project, after an XPG image or an MML transcription is opened, the transcriber will use our Transcription Assistant to transcribe all or part of the document.

Once each word has been boxed, the user can begin the actual process of transcription. Untranscribed boxes are red. The next box to be transcribed can be clicked on, revealing a text field where the transcription can be typed. To emulate the functions of a word processor, the user can move to the next box by simply hitting the space bar in between boxes. Once a transcription has been typed in, its box will turn green and its translation will be entered into the transcription MML and will appear in its exact relative location in the preview window.

The current version of TA allows the user to right-click on a box to annotate the transcription. A primary form of annotation has to do with differentiating graphics or symbols from text. If a box is tagged as an "image" or "symbol" a cropped piece of the manuscript will be copied into the preview window (and into the underlying MML). Currently, we provide the following other types of annotations for text boxes: (i) manuscript annotations such as for stricken text or corrected text; (ii) tagging of abbreviations; (iii) identification of numbers; and (iv) identification of handwriting changes (different author). We foresee adding more of these annotations – such as the tagging of currency and marginalia – as well as second-order tags to identify proper names, names of places, professions, dates and the like.

The one-to-one pairing of wordboxes-to-text transcriptions is broken only by abbreviations, wherein a single word box can be exploded into more than one transcription word. In any case, this pervasive one-to-one correspondence allows the accrual of handwriting-recognition capabilities, which are planned for future versions of the software. We foresee that when users will experience difficulties in transcribing a specific word, they will be able to ask for help by hitting a help key. Using the manuscript metadata for a bounded search – limited to manuscripts that are likely to have the "same hand" – the system will be able to pattern-match the handwriting in the box where the user is having trouble, with a storehouse of boxes from previously completed transcriptions from the same source, yielding transcription suggestions ranked by their different levels of matching. The user will thus be able to pick the suggestion that best fits the sentence being transcribed. We plan to explore the possibility of making this advanced capability available "for a fee" in order to fuel our incentive program and make the inevitable system upgrades self-sustainable over time. When users submit completed transcriptions they accrue internal credits that they can later "spend" to "buy" services like this "transcription help".

Partial as well as completed transcriptions are automatically stored in the UScript Server. Similar to Google Docs, the user can share the new transcription with others or keep it private for a period of time. Privately held transcriptions will automatically revert back to public domain after a selectable number of days, up to a maximum of 2 months. Additional time can be "purchased" using banked credit or via a

"premium" subscription, which will provide funds to maintain and enhance Uscript.

Manuscript with edited text boxes

XML View

<box type="text" corner1x="1" corner1y="1" corner2x="4" corner2y="3"><bold>JHS </bold></box> <box type="text" corner1x="4" corner1y="2" corner2x="8" corner2y="3">Maria </box> <box type="text" corner1x="8" corner1y="2" corner2x="16" corner2y="3">Joseph 1 </box> <box type="text" corner1x="2" corner1y="3" corner2x="4" corner2y="6">El </box> <box type="text" corner1x="4" corner1y="4" corner2x="8" corner2y="6">Diuino </box> <box type="text" corner1x="2" corner1y="3" corner2x="4" corner2y="6"> Caçador </box> <box type="text" corner1x="1" corner1y="6" corner2x="9" corner2y="16"> <Del type="overtrike">La fiera de los montes </del></box> <box type="text" corner1x="2" corner1y="8" corner2x="8" corner2y="21"> Auto </box> <box type="text" corner1x="8" corner1y="8" corner2x="19" corner2y="21"> Sacramental </box> <box type="text" corner1x="6" corner1y="11" corner2x="17" corner2y="14"> Personajes</box> <box type="image" corner1x="16" corner1y="8" corner2x="21" corner2y="14"> …/image1.jpg</box>

Digital Transcription

JHS Maria Joseph

El diuino caçador

~~La fiera de los montes~~

Auto sacramental

XML View

```
<box type="text" corner1x="1" corner1y="1" corner2x="4" corner2y="3"><bold>JHS
</bold></box>
<box type="text" corner1x="4" corner1y="2" corner2x="8" corner2y="3">Maria
</box>
<box type="text" corner1x="8" corner1y="2" corner2x="16" corner2y="3">Joseph 1
</box>
<box type="text" corner1x="2" corner1y="3" corner2x="4" corner2y="6">El </box>
<box type="text" corner1x="4" corner1y="4" corner2x="8" corner2y="6">Diuino
</box>
<box type="text" corner1x="2" corner1y="3" corner2x="4" corner2y="6"> Caçador
</box>
<box type="text" corner1x="1" corner1y="6" corner2x="9" corner2y="16"> <Del
type="overtrike">La fiera de los montes </del></box>
<box type="text" corner1x="2" corner1y="8" corner2x="8" corner2y="21"> Auto
</box>
<box type="text" corner1x="8" corner1y="8" corner2x="19" corner2y="21">
Sacramental </box>
<box type="text" corner1x="6" corner1y="11" corner2x="17" corner2y="14">
Personajes</box>
<box type="image" corner1x="16" corner1y="8" corner2x="21" corner2y="14">
.../image1.jpg</box>
```

Digital Transcription

JHS Maria Joseph

El diuino caçador

~~La fiera de los montes~~

Auto sacramental

Personajes

Above: Illustration of how the Transcription Assistant (TA) will work

Contribution Accountant

The Contribution Accountant (CA) is an opensource web application that runs on a Linux machine running an Apache server with a mySQL database backend. This may or may not reside on the same server as the Archive Assistant (AA). This application is currently only in the planning stages. The CA will

provide a set of APIs through which a number of account services will be accessible to other applications. It is unlikely that individual users will have direct access to the CA, since it is probably going to be an internal, backend application supporting the AA and TA.

We foresee an unobtrusive and semitransparent accounting system that will keep track of the contributions made by each transcriber in order to attribute appropriate credit to the authors that submit new or improved transcriptions. Such contributions could be rewarded by a sort of "credit" that can be used to acquire advanced services from our system as discussed above. Transcribers would also be rewarded through an automatic academic citation system that recognizes an author's contribution as bona-fide intellectual property which must be properly referenced whenever an author's transcription is used in a project or cited in a paper.

The credit system is predicated on the (free) registration of each user and on the system's ability to unequivocally recognize legitimate users through passwords. After registering, users will receive a virtual "bank account" where their credits will be posted by the CA. The essence of the CA accounting system is based on before-and-after comparisons between what the AA sent out to the user and what comes back from the user for each manuscript page. A differential engine will quantify the number of changes made to the original transcription as well as to the word boxes in the manuscript image. The net change will be the basis for the credit awarded to the user. However, to avoid spurious submissions from users bent on "gaming" the system, the initial credit would be awarded on a provisional basis as a percentage of the total earned by the transaction. The "discounted" rate of the temporary credit will be based on the credit-worthiness of the user, which depends on his or her credibility, which is accrued over time, by indirect peer review.

Peers will in essence give their approval by implicitly confirming the quality of submitted changes, simply by not putting forward additional corrections to what the original contributor(s) had provided. A sort of tacit consent will thus enhance the credibility of users whose submissions pass muster with subsequent users. In addition to this process of "dynamic accreditation", some archives may choose to start from an official list of known, fully accredited professionals to expedite the "reputation management" component of the system, using a form of "a priori accreditation"

Our aim is to arrive at a universal registration system that will maintain the distributed and emergent nature of the overall system without compromising the quality of the transcriptions that bubble up to the top of the heap after successive refinements by a variety of contributors.

Users could accumulate "transcription credits" and "word box credits" (possibly on different "pay scales") when they make contributions to the system. Credits could in turn be usable to "pay" for such services as ondemand scanning of manuscripts, handwriting recognition assistance, remote storage of project files on the archive servers, automatic transcription processing, advanced searching, and others.

Work plan

Currently, only the Transcription Assistant has been developed to the point that it can be used commercially by end users. The Archive Assistant needs work on the backend software to function as required, and the Contribution Accountant has yet to be implemented. Once these three parallel components are in place, Uscript will be ready for testing by the archives collaborating with us on this project.

The attached budget includes the onetime cost of purchasing a server on which we will store software code and manage the websites related to the Uscript project. We have also requested, on a

recurring annual basis, funds to cover the cost of domain names and high-traffic, high-bandwidth web hosting for what will become a heavily used website.

We estimate that Uscript can be fully developed and made publicly accessible in three years, especially given the work already done, both in conceptualization and concrete programming, on the project to date. An outline of our objectives for this proposal is below, the details of which will be given in the following sections:

Year One

- Review, organization and validation of code base
- Revamped frontend CMS
- Developer blog and forums
- Bug reporting and support system (tickets, etc.)
- Adding a throbber/loading bar
- Release of core

Year Two

- Browser history support
- User projects
- Splitting/merging boxes
- User forums
- User management
- Incentive system
- Straight-to-text
- Importing options
- Submission incentives

Year Three

- Expanding support for other languages and other archives
- User pages
- Metadata management and searching
- Enhanced optical character recognition (OCR)
- Algorithm optimization

Project management and structural improvements

The Uscript project cannot really take off until it is fully organized as a real open source initiative, with all of the infrastructure that makes such enterprises flourish. These first objectives will legitimize Uscript as a true open-source project, open to the world and workable by those who want to continue to the code development, as well as usable by those who want to begin using Uscript for the development of digital transcriptions of ancient manuscripts.

Revamped frontend content management system (CMS) year one

The open-source project itself should have a well thought-out front page, starting from the existing site organized as a content management system (CMS) located at http://uscript.veniceprojectcenter.org.

The [current SourceForge repository](#) and site also needs to be made more generally appealing along the lines of other major open source efforts.

Review, organization and validation of code base year one

The Subversion code repository should be reviewed and reorganized for release in order to facilitate open source development. Module validation procedures should be developed and applied to existing code to prepare for core release.

Developer blog and forums year one

This idea list should itself become a "live" document to expand its reach and availability to the open source community. In parallel, a Uscript blog and one or more forums should be created to allow developer interactions and to foster the birth of a real Uscript community.

Bug reporting and support system (tickets, etc.) year one

To prepare for core release, we will need to put in place a bugreporting/ticketing system with alerts and follow ups so that users can receive a reasonable response to issues that will emerge once code is released.

Release of core year one

The main goal for year one is to release a core system that is fully operational, albeit limited in functionality, and is supported by a full-fledged open source infrastructure ready to leverage developers around the world and ready to support pioneering users with adequate resources and response-times.

Transcription Assistant enhancements

Depending on the level of interest in the project, we will continue also to develop the core modules, starting with the Transcription Assistant, which may well represent the bulk of the initial core release. Starting with the existing application, the following enhancements could be made:

Adding a throbber/loading bar year one

In the current implementation of the Transcription Assistant, there is no way to monitor the status of any ongoing loading or processing. As a result, users waiting for the program to load a large element or finish an intensive task are left without any indication that the application is still actually functioning. This is addressed in browsers with elements like loading bars and throbbers.

Throbbers are icons located typically in the top-right corner of a browser's screen which run a simple animation while the browser is busy. This is a simple method for reassuring the user that all is well and prevent them from mistakenly closing out the application. While our current state of implementation does not necessitate a throbber or a loading bar, almost any extension of functionality would be greatly aided by the presence of such a device.

Browser history support year two

One of the Google Web Toolkit's (GWT) commonly touted features is its browser history support. Since the browser history is tied to interpreted HTML, JavaScript developers often do not or cannot utilize it. That means that when users move between pages in such applications, they cannot rely upon the browser's built-in navigation tools. GWT allows the creation of new history items that can be placed on a browser's history stack. While this is not a core feature, it does not appear prohibitively difficult to

implement compared to its benefits in usability.

User projects year two

Realistically, users are not going to simply look through one or two documents for the whole of their research. Instead, they are going to collect lists of documents that they reference, and it would be very useful for the Transcription Assistant to facilitate this. The idea of a project, then, would be to maintain a collection of links to relevant files. Since the project file is simply a list of references, it could easily be exportable to and importable from local files, and could easily be shared by multiple users.

Splitting/merging boxes year two

There is a problematic case in emergent transcription where there is disagreement over whether a section of text is one or multiple words. Worse still, such contested sections may overlap and be interwoven. Our proposed solution would be to create tools for splitting and merging boxes. These splits and merges would themselves need to be suggestions, and as such, would be stored separately from the individual suggestions. Essentially a section that was originally marked as two words could be merged together, creating a third suggestion-storing object for when the boxes were viewed as one. The user would then need some visual indication in order to switch between the two views of the wording. Splitting would work similarly in the other direction.

Straight-to-text year two

A straight-to-text tool would be fairly simple to write, and is very necessary in the next release of the software. This tool would take the top ranked suggestion for each word (or the word that the current user has chosen) and output the words in sequence (as defined by the user with the help of the auto-ordering algorithm). The output could go to an output file to be saved locally, or to a secondary "Preview" tab.

Expanding support for other languages and other archives year three

Currently, our project only has a language entry present (for English), and has no specialization for different archives or formats. Since we would like to see this project implemented in various countries by different archives, it will be important for future groups to spend time working on implementing new language support and making metadata friendly to differing formats. The current application already contains the groundwork for implementing internationalization. Also, thankfully, most archives primarily use an item reference number that contains all of the information they need, but the accommodation of changes such as specialized forms must nevertheless be anticipated.

User Experience and Management

Once the core Transcription Assistant is released, we will need to begin managing users, so we should plan to release tools to support them in parallel with the initial core release.

User forums year two

Another useful addition to the Uscript system that future work should implement would be a set of user forums. These would allow interaction between users, beyond simply assisting one another in transcribing manuscripts, by way of offering suggested transcriptions and giving an thumbs-up-or-down vote on existing suggestions. Users would be able to have more open discussions with one another about transcribing and the Uscript system in general. Along with providing a much stronger sense of community and cooperation, user forums could serve as a mechanism with which to collect feedback

about the usability of the system and suggestions for improvement.

User management year two

Another feature that would be nice to see in future iterations of the Transcription Assistant is user management. The possibility of malicious users working within the system should be recognized as a potential problem and some protection scheme put in place to combat behavior that is clearly subversive. One simple way this could be accomplished is by allowing users to report other users who they suspect are causing trouble. Reputation and credibility management also present an opportunity both to improve the usability of the Transcription Assistant and help offset the efforts of malicious users. While a simple reputation system exists in the current system, a future team should take the time to devise a more sophisticated scheme.

Incentive system year two

One of the long-term goals stated early on in the initiative was the desire for some sort of incentive program for system users. This could include but is not limited to monetary credit for providing transcriptions, to be funded by yearly service subscription fees. This could also be pushed along by providing higher credit to users with higher transcription reputations, and by charging credit for access for advanced features such as an optical character recognition algorithm or for access to view a transcription.

User pages year three

Many well-established, successful websites have user pages that act as a sort of customized headquarters to make browsing that site easier for each user. A user homepage on the Uscript website should list transcriptions they have worked on recently, a user's projects, transcriptions that are similar in content to manuscripts that user has worked on, and similar transcriptions listed by era or location.

Overall system improvements

Together with specific functions, more "academic" and "research" issues will also need to be addressed to improve the system's performance.

Metadata management and searching year three

The current Uscript system provides the framework for taking a set of metadata obtained from (e.g.) the Venice State Archive and allowing users to search for existing transcriptions with those parameters. Ideally, users will be able to seek out manuscripts/transcriptions on a given topic by searching the body of text in a transcription in addition to manuscript metadata. This will increase the amount of related material users are able to discover to further their research. We intend to explore adopting D-space to manage this aspect of the project, as well as the addition of a metadata and content (i.e., transcribed text) search feature.

Enhanced optical character recognition (OCR) year three

Enhanced optical character recognition has long been a goal of the initiative. As it currently stands, the image is analyzed in order to determine where individual words lie within the page ("auto-boxing"). This is commonly the first step taken in modern OCR algorithms. We would like to see this taken a step further into analyzing the content within the boxes. (As a side note, the success of this step would be dependent on a reliable auto-boxing of the image.)

The method for analysis is yet to be researched, but some ideas have been proposed. Typical methods

are unlikely to succeed in for this project, as the intricacy and variability of manuscript text varies so widely. It wouldn't be appropriate to compare the word images to any standard set (or sets) of data in order to discover individual letters and proceed to build words. A more feasible option might be to utilize previously transcribed words, paired with some sort of quantitative analysis of the digital image.

If, once the words on a page were boxed, the image data within each box could be extricated and stored, individual words could be compared to each other. If patterns matched, the system could use the user recommended transcriptions for one word in order to make suggestions for the other unknown word. It would be an interesting algorithmic problem to solve in how the software could analyze, quantify, and compare each word image.

It would be extremely computationally intensive to compare, pixel by pixel, one image to another, and would also be very difficult to determine whether it was a valid "match". One possible solution that has been raised is the possibility of quantifying the pen strokes within a word image to make for easier pattern matching. Other possible optimizations of this algorithm might take into account the metadata for the manuscript that a word is taken from, such as time period, author, or language. Clearly, a match would be more likely with word images that are known to be from the same author, written in the same hand.

Importing Options year two

Expanding options for importing a manuscript into the system through the Archive Assistant is a necessary improvement. When importing a manuscript, the archivist should be required to enter a specific set of metadata about the manuscript. A preliminary boxing of each page of the manuscript should also be required. There should be some options presented to the archivist about how the documents should be boxed. Some options might include manually boxing the page(s), manually setting threshold values for the auto-boxing algorithm, displaying some sample manuscript images with preset threshold values for similar manuscripts, or the option to manually box a single page of a multipage manuscript in order to have the genetic algorithm find optimal threshold values for the rest of the pages.

Submission Incentives year two

One of the long-term goals stated early on in the initiative was the desire for some sort of incentive program for system users. This could include but is not limited to monetary credit for providing transcriptions, to be funded by yearly service subscription fees. This could also be enhanced by giving more credit to users with higher transcription reputations, and by charging users for access to advanced features such as an optical character recognition algorithm or to view a transcription. None of these incentives should interfere with Uscript's free and open-source nature.

Algorithm Optimization year three

Most of the server-side processes have somewhat complicated algorithms, and most of them can and should be optimized to run more efficiently. The auto-boxing algorithm uses a smearing algorithm that may not be the most accurate choice. The auto-lining algorithm could be written to run more efficiently. The genetic algorithm was written to run infinitely, or until halted by a user. Realistically, it needs to have some way to decide when it has found a satisfactory solution. The inner workings of the genetic algorithm and auto-boxing are discussed in depth in previous reports, and the current auto-ordering algorithm is explained within the code.

Staff Fabio Carrera, PI:

Director of WPI Emergent Systems Lab. Former director of the Laboratory for Humanistic Informatics at the University of Venice. Instructor of Computer Applications in the Humanities at various University of Venice Computer Centers. Developed several historical databases to support historical journal articles and books including Alla Luce della Luna (author S. Piasentini) and The Venetian Money Market 1200-1500 (author R.C. Mueller). Initiator of the Uscript concept and advisor of all undergraduate projects to date. Will supervise the entire project and coordinate academic research, software development and dissemination.

Stephen Guerin, lead consultant:

President of Santa Fe Complex. Instructor at Santa Fe Institute Summer School. Expert in complexity theory and applied complexity. Primary consultant for the implementation of the Uscript system. Will oversee the professional programming and the release and operation of the server and online systems.

Will help implement a search and correlation system similar to the one he created for the Santa Fe Institute library (http://www.redfish.com/projects/SFISemanticWeb/SFIGraph.html)

Stanley Selkow, advisor:

Computer Science professor, WPI. Coadvisor of all Uscript research projects to date. Will continue to advise academic projects on the topic.

Gary Pollice, advisor:

Computer Science Professor, WPI. Experienced advisor of Open Source projects. Will coadvise student projects and provide consulting on technical issues.

Dissemination

Our work on Uscript will be made available to the public via the Internet, including all source code, documentation as well all of the content. As per NEH's requirements, we will also release a white paper detailing our work, solutions found to problems encountered, programming techniques used, and other findings. We also plan to publish our results in appropriate journals and conferences (ICDAR), as we have already done in the past.

In the end, we plan on unleashing this self-propagating emergent system as open-source software onto the Internet by the end of 2013, so that the entire system, including the three applications themselves, will grow on its own without central control. If successful, this system could really "make history" by exponentially expediting the production of historical research in the whole world.

Continued development past end of NEH support

After support from this proposal ceases at the end of year three, Uscript will be in place on a production server, and in use at several major archival collections in New England, the rest of U.S. and at the Venice State Archives in Italy. We expect to have attracted and retained a community of open-source developers and programmers over the course of the project, and one need only examine the cases of open--source "killer apps" like Linux, Mozilla Firefox, and MediaWiki to conclude that users stay attached to projects they find useful and interesting. We will structure the development of Uscript in such a manner that it, too, will be a project that developers want to keep working on.

WPI has remained dedicated to Uscript for over a decade. We expect to see continued contributions to the project from WPI after the end of NEH funding. There will be annual costs associated with the project, arising from the need for centralized server space and backups, system administration, and the coordination of ongoing software development. We intend to request additional funding from the Mellon foundation to support the ongoing efforts to make the system self-sustaining.

We also plan to explore clever financing schemes to make the project self-supporting, associated with the possible fee-for-service functions that we will be investigating in the context of Uscript's core functionalities.

Technical details

We will be using the following hardware/software configurations in the course of our project:

Hardware: Development Server Quad Core, e.g., Dell PowerEdge or Apple XServer **Software:** Server support and subscriptions **Database:** mySQL or PostGreSQL **Data processing:** OCR, scanning, conversion **Image resolution, compression, and enhancement:** max allowable **Tagging schema/markup:** Dublin Core and MARC standards **Documentation:** JavaDOC or similar **Aids to discovery:** Complexity algorithms, such as those developed at the Santa Fe Complex (http://www.redfish.com/projects/SFISemanticWeb/SFIGraph.html) **Future preservation and migration plans:** We will migrate Uscript to a production server in year 3

Budget Justification

Salary and Wages

Two months of the PI's summer salary are requested for each year of the three-year proposal. The salary rate is based on the PI's current appointed academic year salary rate, with an annual raise of 4% applied in years two and three.

Furthermore, we are requesting funds to support the work of approximately five undergraduate students every year. The funds will support one Computer Science research project every year, to investigate the thorniest technical issues before implementation (approx. $5,000/year). Remaining funds (~ $10,000/yr) will support work-study undergraduates during the academic year as well as summer interns at the Santa Fe Complex. This stipend rate is based on WPI's established rates for fiscal year

2010.

Consultant Fees

Because this project relies heavily on computer programming, we intend to employ professionals at key stages. Open-source programming, as will be used in our project, typically relies on the time and energy of volunteers. However, to ensure that Uscript stays on schedule, and to provide dedicated assistance as needed, we have requested funds to permit us to hire professional programmers to release the core elements of the system. The daily rate adopted is $400/day for 90 days of programming each year of the project.

Travel

We have included a budget for one trip by the PI from Boston to Santa Fe and for one for the primary consultant (Steve Guerin) from Santa Fe to Boston every year, with a stay of five days each time. Moreover, we have included a trip for two people to Washington every year to report and debrief with NEH project officers.

Other Costs

Every year, we have budgeted $500 for web hosting and domain name services. In year one, we have also included $5,000 for the purchase of a development server dedicated to the project.

Appendix C: Scheda dello Scheletro

129

scheda dello scheletro

numero di elemento della tomba

in sarcofago?

reperti con lo scheletro?

in connessione?   completo?   disturbato?

posizione di:
cranio

braccia

gambe

osservazioni
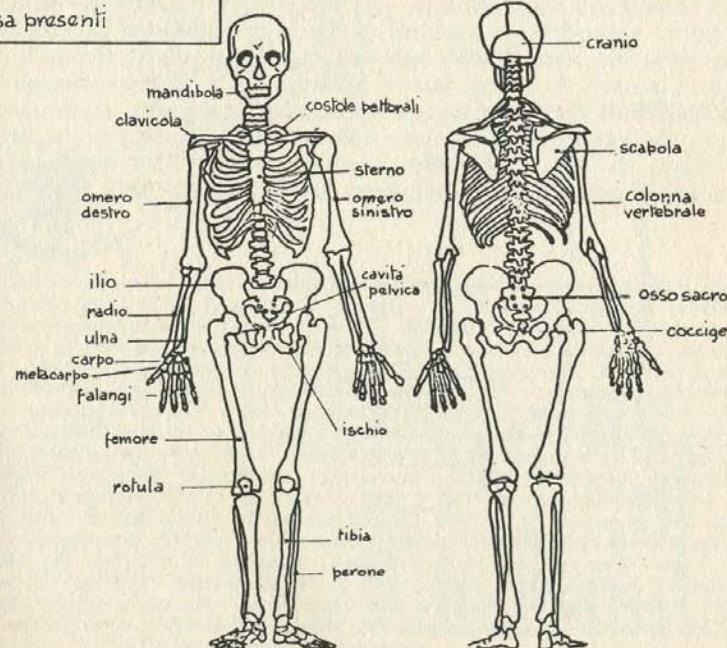
condizioni delle ossa

ossa presenti

mandibola
clavicola
costole pettorali
sterno
omero destro
omero sinistro
ilio
cavità pelvica
radio
ulna
carpo
metacarpo
falangi
ischio
femore
rotula
tibia
perone

cranio
scapola
colonna vertebrale
osso sacro
coccige

Appendix D: Sa per La Classificazione di Materiali Organici
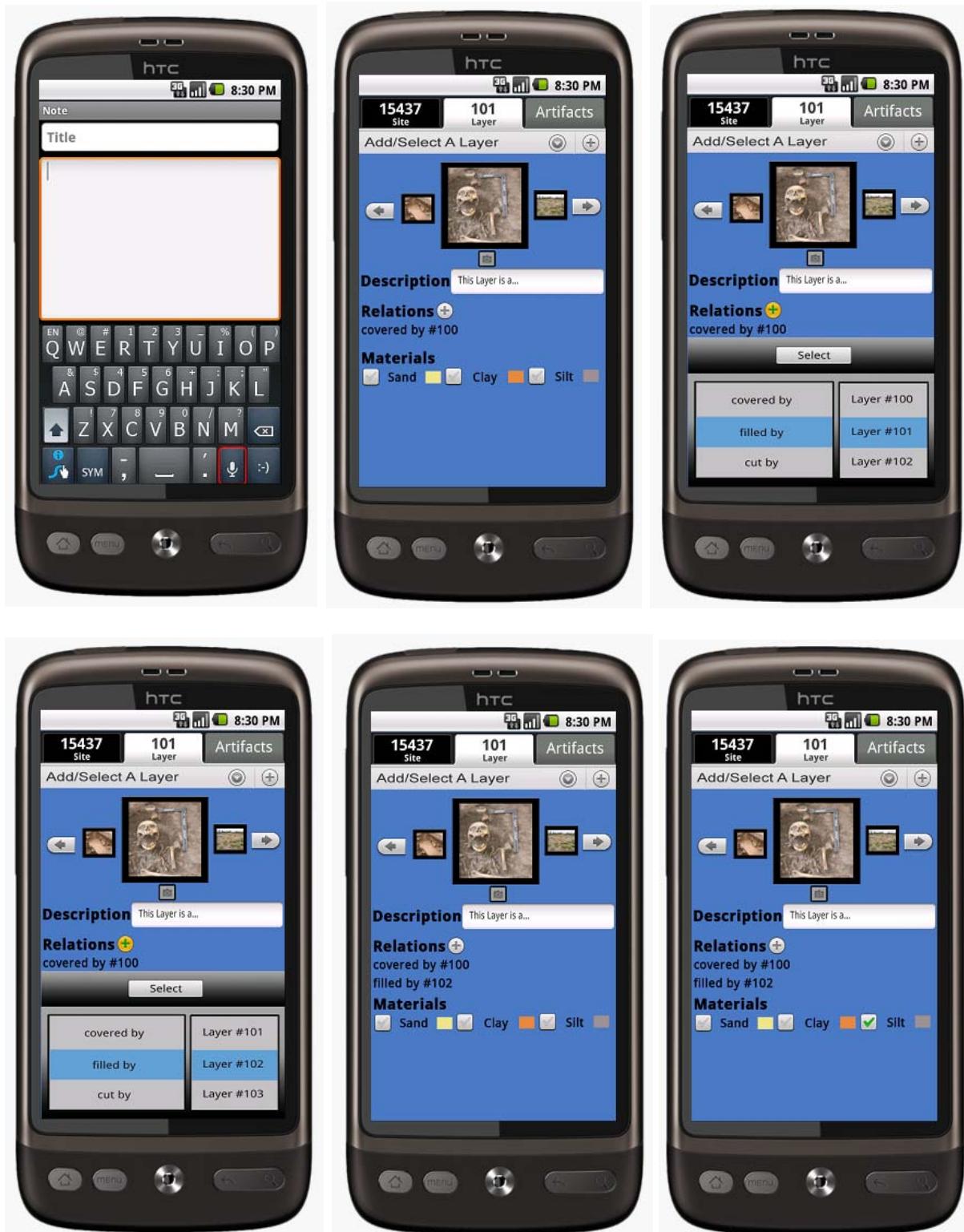
SCHEDA PER LA CLASSIFICAZIONE DI MATERIALI ORGANICI	
MINISTERO PER I BENI E LE ATTIVITA' CULTURALI	**DR. MARCO BORTOLETTO** Via L.Bissolati 6/7 30172 Mestre Ve P. Iva 03073310272

Località:				Scheda n°	
Sigla:		Campata:		Vano:	
US:		n° reperti		Scheda cumulativa:	
Materiale:		Epoca:			
Forma/oggetto:					
Classe:					
Tipo:					

Descrizione:

n° inventario	Foto
n° disegno:	
n° foto col:	
n° negativo col:	
n° foto b/n:	
n° foto b/n:	
n° foto digitale:	
note:	

Resposabile dello scavo:	Fotografo:
Direzione tecnica:	Compilatore della scheda:_

Appendix E: Layer Form and Tomb Form Index

	ELENCO US-USM	Sigla	MUR FDM 33	Anno	2008		p.	elenco		1
US	Descrizione	Posizione	coperto da	tagliato da	riempito da	gli si appoggia	copre	taglia	riempie	si appog-gia a
100							101			
101			100				102,104,112			
102			101,112				103			
103			102				104			
104			101,103				105			
105			104				106			
106			105				107			
107			106				108			
108			107				109			
109			108				110			
110			109				111			
111			110							
112			101				102			

Appendix F: Archaeological Application Screen Shots
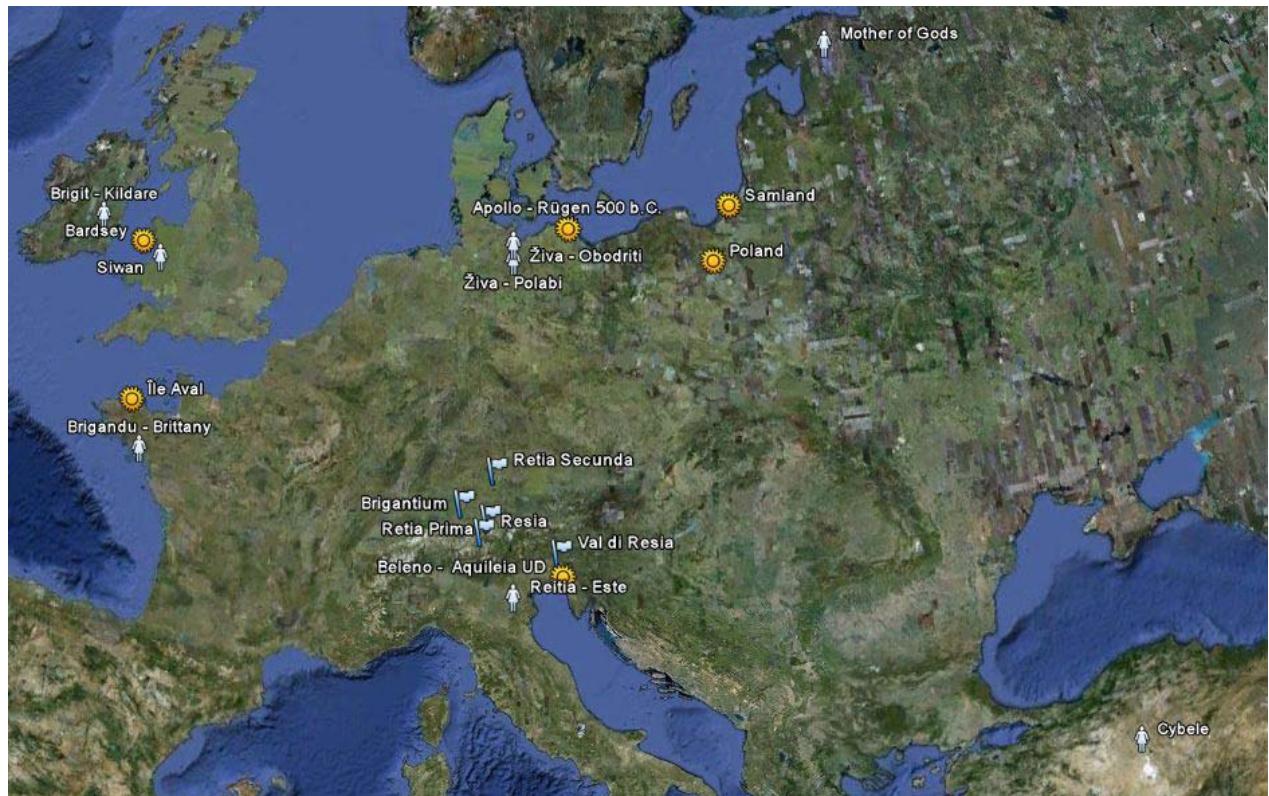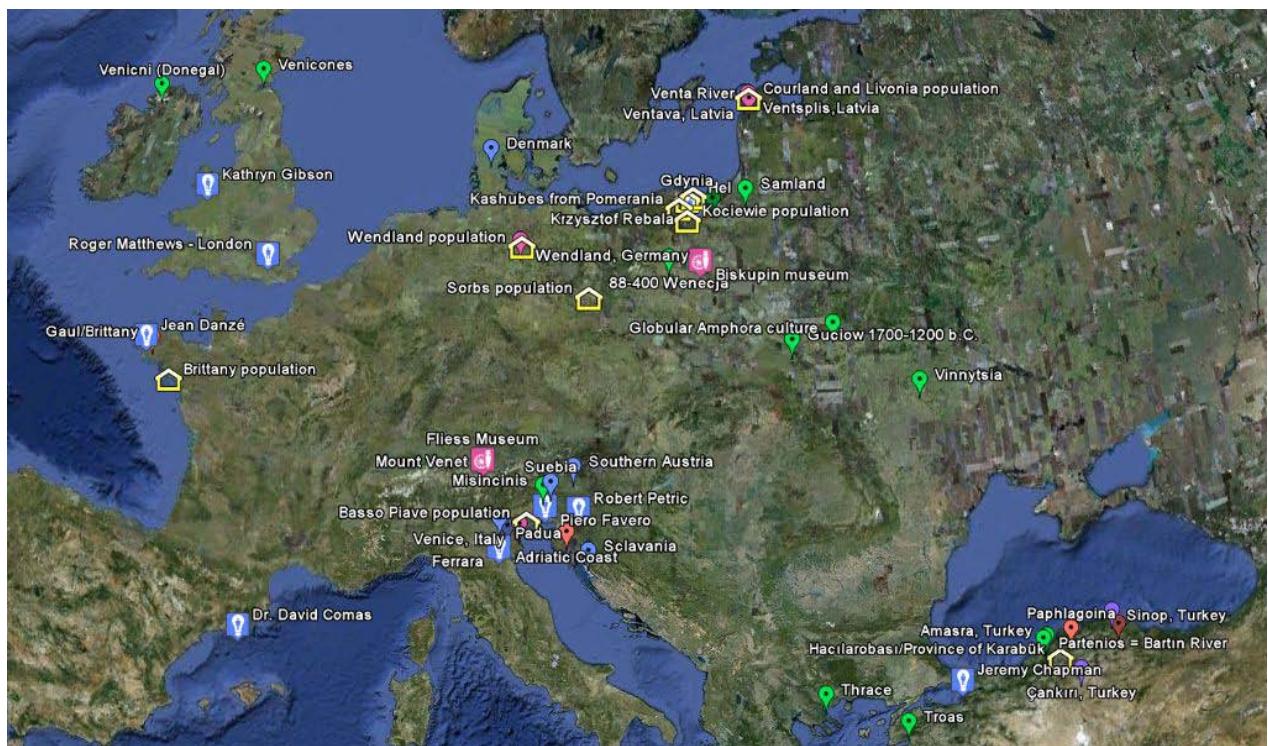
Appendix G: Our Google Maps
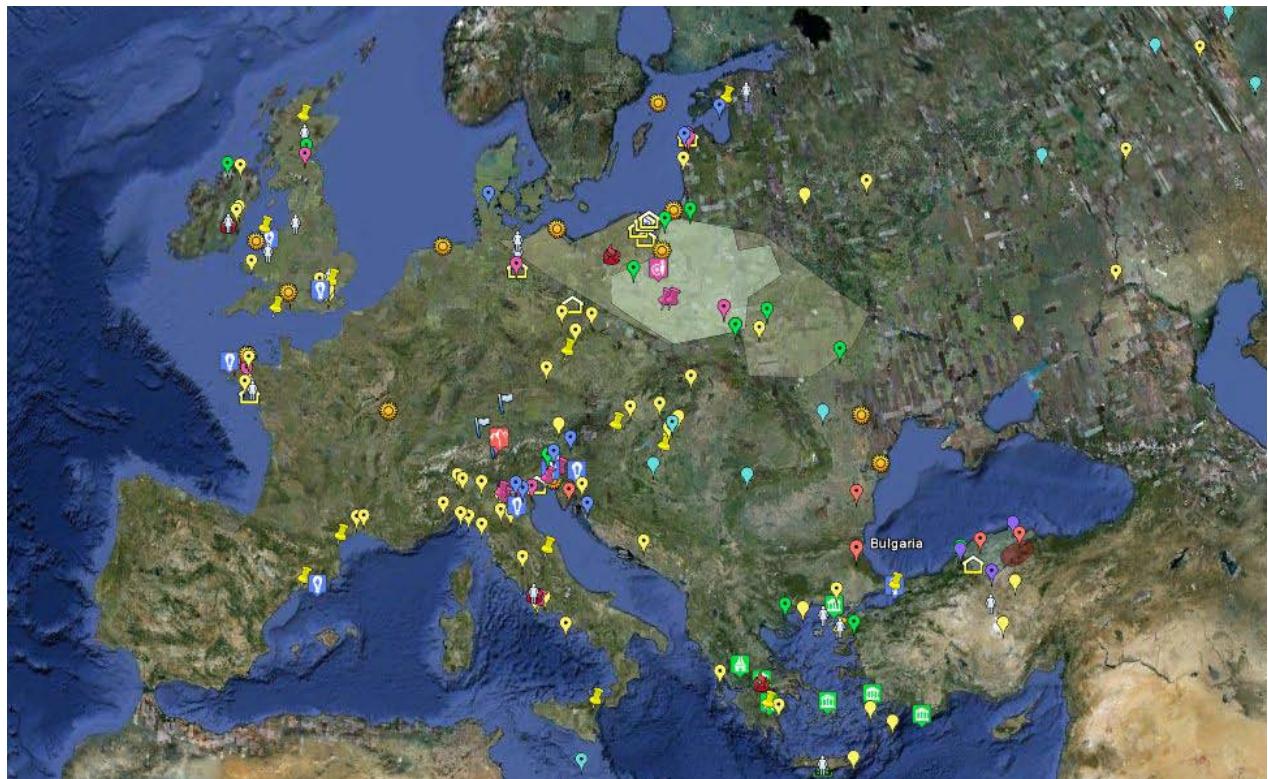
Urnfield Culture Map

Mythology Map
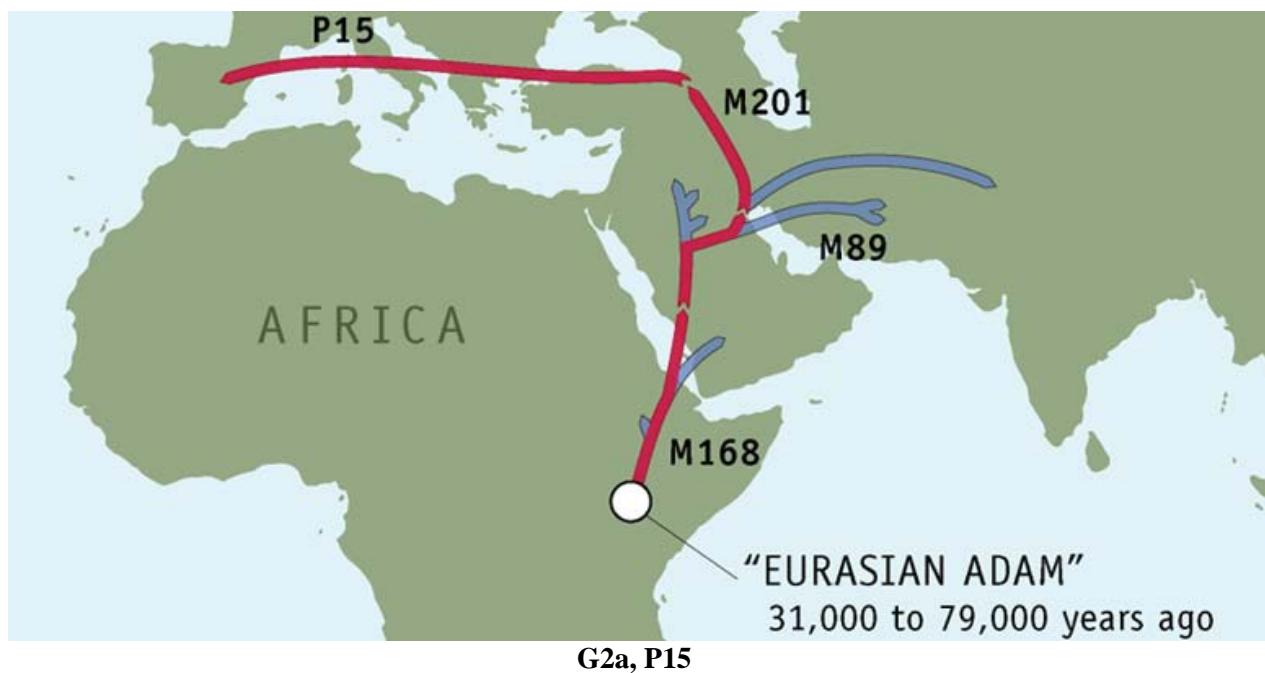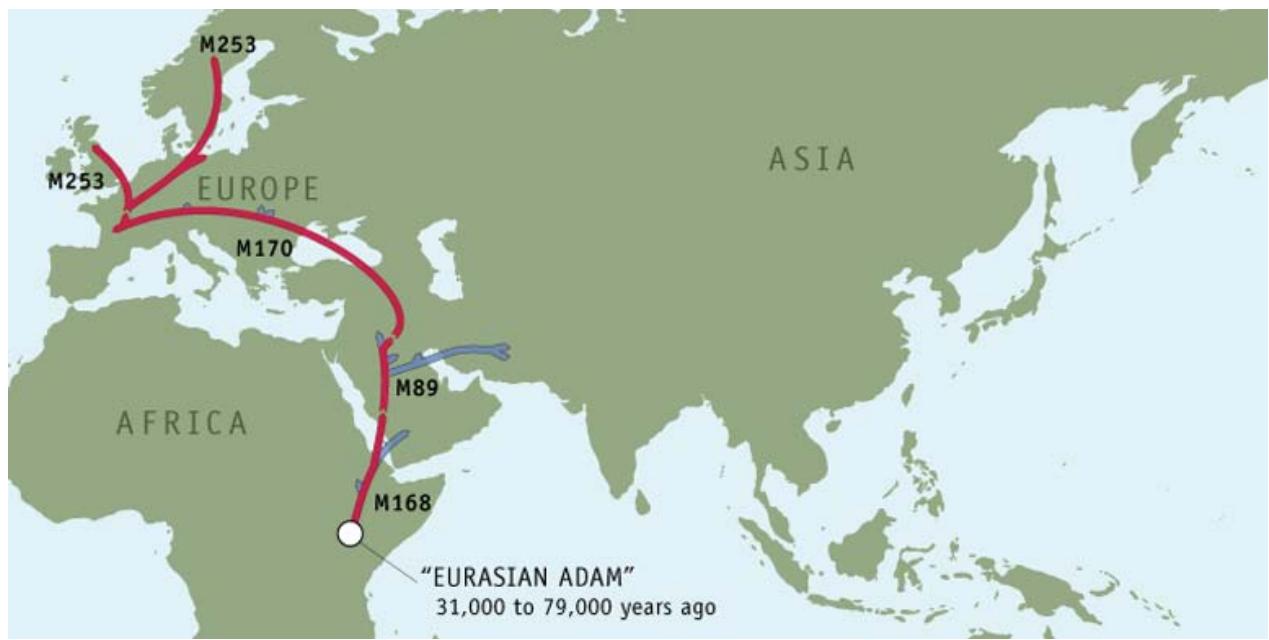
Our Map

Collaborative Final Map

Appendix H: Genographic Project Haplogroup Maps

E1b1b1, M35

G2a, P15

I1, M253

I2a, P37.2

I2b1

J2, M172

J2

L

R1a1

R1b

T, M70
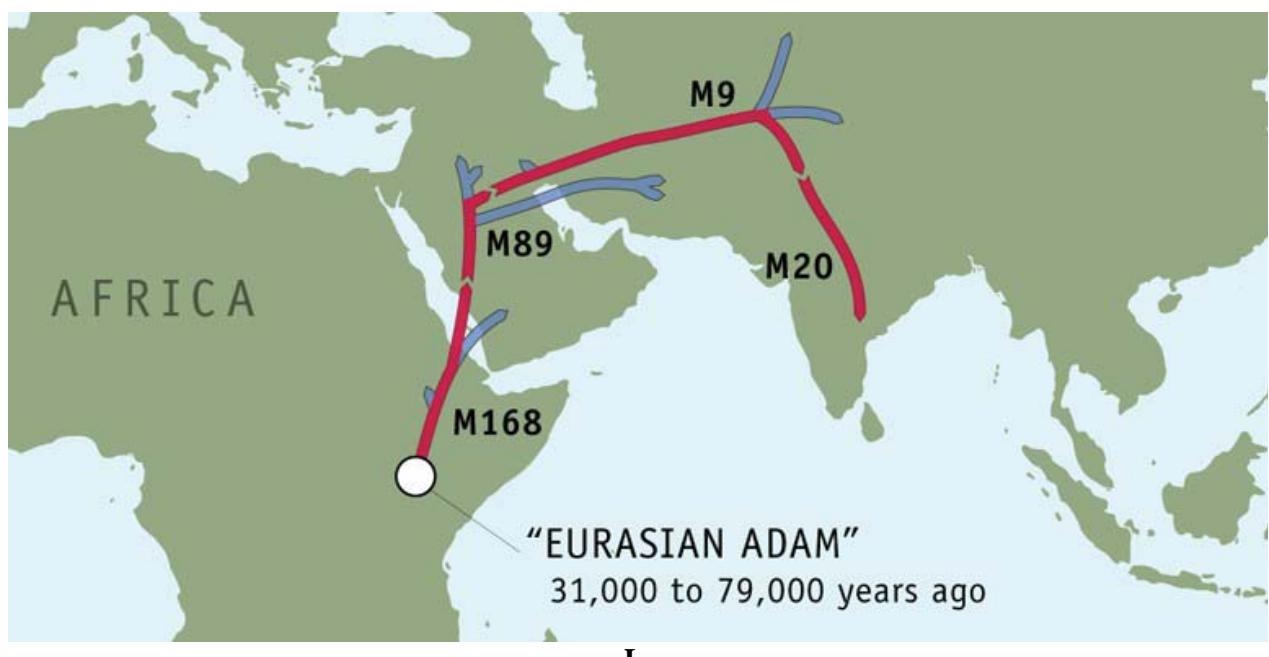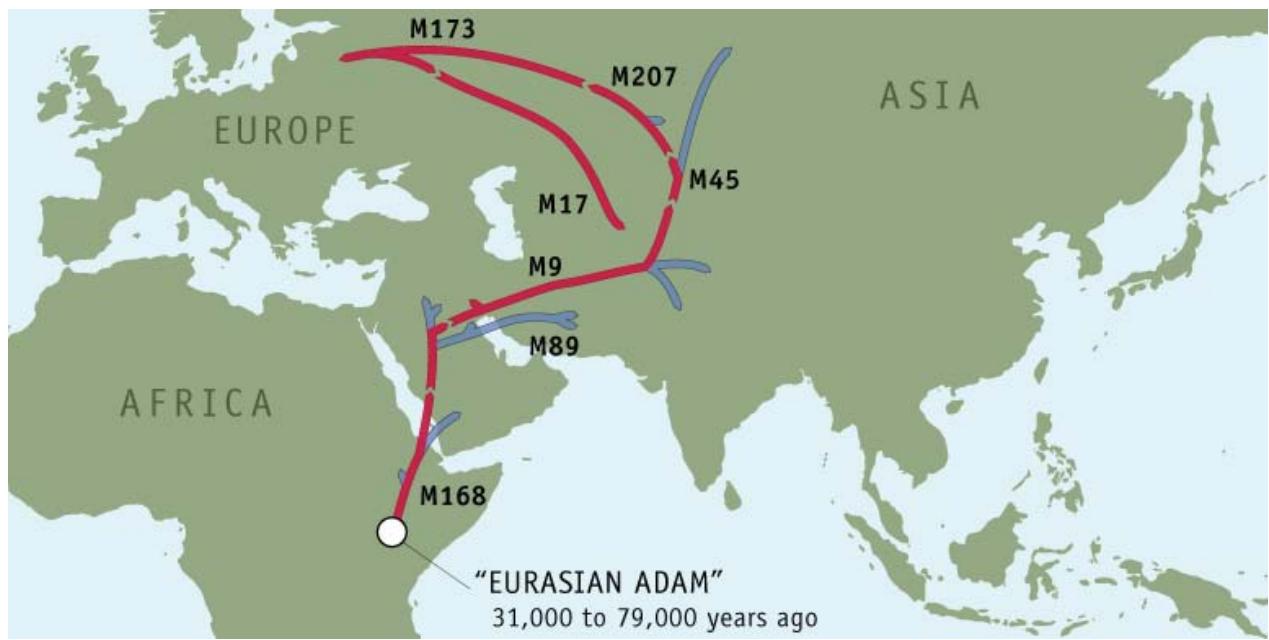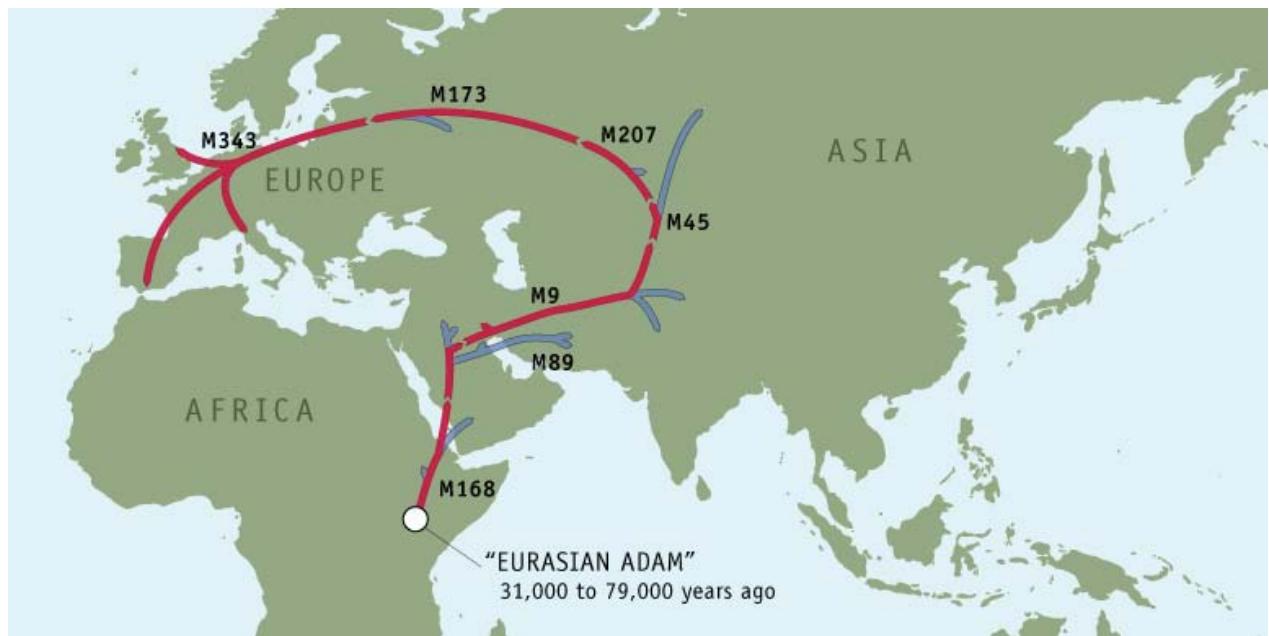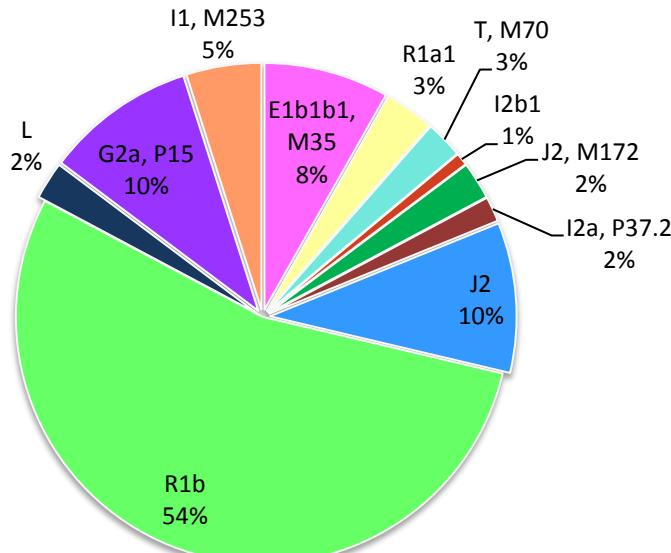
| Local ID	Country	State (Region)	District (Province)	City	Maternal GM Place of Birth	Maternal GF Place of Birth	Paternal GM Place of Birth	Paternal GF Place of Birth	Not rec'd	E1b1b1, M35	R1a1	T, M70	I2b1	J2, M172	I2a, P37.2	J2	R1b	L	G2a, P15	I1, M253
										032	048	064	070	037-3	063	037	050	039	060	62
VE-01	Italy	Veneto	Venice	Venice	Milan	Milan	Milan	Venezia								X				
VE-02	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Robbio									X			
VE-03	Italy	Veneto	Venice	Venice	Cheoggia	Cheoggia	Venezia	Venezia		X										
VE-04	Italy	Veneto	Venice	Venice	Venezia	Venezia	Vicenza	Vicenza										X		
VE-10	Italy	Veneto	Venice	Venice	Udine	Venezia	Veneto	Venezia											X	
VE-11	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Veneto									X			
VE-12	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Chioggia								X				
VE-21	Italy	Veneto	Venice	Venice	Murano	Murano	Venezia	Venezia		X										
VE-22	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Venezia									X			
VE-23	Italy	Veneto	Venice	Venice	Chioggie	Chioggie	Venezia	Venezia												X
VE-54	Italy	Veneto	Venice	Venice	Verona	Verona	Bologna	Verona			X									
VE-55	Italy	Veneto	Venice	Venice	Veneto	Veneto	Veneto	Veneto									X			
VE-56	Italy	Veneto	Venice	Venice	Venezia	Milano	Venezia	Venezia											X	
VE-57	Italy	Veneto	Venice	Venice	Veneto	Veneto	Veneto	Veneto									X			
VE-59	Italy	Veneto	Venice	Venice	Treviso	Treviso	Venezia	Venezia								X				
Ve-64	Italy	Veneto	Venice	Venice	Veneto	Veneto	Veneto	Veneto												X
VE-65	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Veneto									X			
VE-66	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Venezia											X	
VE-67	Italy	Veneto	Venice	Venice	Venezia	Venezia	Ravenna	Ravenna								X				
VE-71	Italy	Veneto	Venice	Venice	Padova	Padova	Bassonode (VI)	Bassonode (VI)						X						
VE-72	Italy	Veneto	Venice	Venice	Udine	Udine	Treviso	Venezia												X
VE-85	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Venezia						X						
VE-86	Italy	Veneto	Venice	Venice	Belluno	Belluno	Varsavia	Zara									X			
VE-122	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Venezia											X	
VE-123	Italy	Veneto	Venice	Venice	Treviso	Treviso	Belluno	Venezia							X					
VE-128	Italy	Veneto	Venice	Venice	Treviso	Treviso	Venezia	Venezia							X					
VE-152	Italy	Veneto	Venice	Venice	Venezia	Venezia	Venezia	Venezia											X	
VE-156	Italy	Veneto	Venice	Venice	Feltre (BL)	Padova	Pernomia (PD)	Pernomia (PD)									X			
								TOTALS:	34	10	4	3	1	3	2	12	66	3	12	6
									22%	8%	3%	2%	1%	2%	2%	10%	54%	2%	10%	5%

Haplogroup Distribution in Venice

- I1, M253 5%
- T, M70 3%
- R1a1 3%
- E1b1b1, M35 8%
- I2b1 1%
- J2, M172 2%
- I2a, P37.2 2%
- L 2%
- G2a, P15 10%
- J2 10%
- R1b 54%

Appendix J: Swabbing Instructions and Kit

Whatman Sterile Omni Swab
Instructions For Use

The Whatman Sterile **Omni Swab** is designed for collection of buccal cell samples for DNA testing. To obtain a buccal cell sample, the tip end of the **Omni Swab** is rubbed against the inside of the cheek. Unlike cotton-tipped swabs, the **Omni Swab** is made of absorbent material specifically designed for the collection of buccal cells. The single use swab has a unique removable collection tip to assist in processing the sample. While collecting a buccal sample is quick, easy and painless, it needs to be done correctly.

Suggested Directions for Use
The following directions are provided to illustrate an acceptable collection protocol. Each facility should establish a sample collection protocol that meets the objectives of collection and testing facilities.

Sample collection:
The person providing the buccal cell samples should not eat or drink immediately prior to giving the sample. If food or drink has been recently taken, it is suggested that the mouth be rinsed with water prior to sampling. The person taking the samples should wear biohazard-barrier gloves and avoid contact with the *Omni Swab* collection tip.

1. Open the *Omni Swab* packaging at the handle end and carefully remove swab. Do not touch the collection tip of the swab.

2. Holding the handle end of the *Omni Swab*, scrape the collection tip firmly against the inside of the cheek 5-6 times (about 10 seconds) being careful not to depress the plunger that ejects the tip.

3. After taking the sample, eject the tip by firmly pressing the plunger at the end of the handle into a labeled 2 mL microcentrifuge tube.

4. If desired, repeat sampling procedure with a second swab using the other cheek. Eject tip into a different labeled 2 mL microcentrifuge tube.

Storage:
After sample collection, tip can be kept at room temperature when processed immediately. If storage is necessary, freeze tips at -20°C.

DNA Extraction:
DNA can be extracted from the tip using standard laboratory procedures. DNA yields of 500-2,000 ng DNA are typical. Actual DNA yield will vary depending on the original DNA concentration, collection efficiency and extraction procedure.

Caution: Read instructions prior to use.
Single use only. Do not use if package has been previously opened or damaged. Federal law (USA) restricts this device to sale by or on the order of a physician.

To obtain Whatman® Sterile **Omni Swabs**, Catalog #WB10-0004: Contact Whatman® toll free at 1-866-787-3362.

Whatman®
www.whatman.com

Appendix K: Worcester Polytechnic Institute IRB Form

WORCESTER POLYTECHNIC INSTITUTE
Institutional Review Board
Application for Approval to Use Human Subjects in Research

WPI IRB use only
IRB # _____
Date: _____

	WPI IRB use only
This application is for: *(Please check one)* ☒ Expedited Review ☐ Full Review	

Principal Investigator (PI) or Project Faculty Advisor: *(NOT a student or fellow; must be a WPI employee)*

Name: Fabio Carrera   Tel No: 508-615-5333   E-Mail Address: carrera@wpi.edu   ☐

Department: Interdisciplinary and Global Studies Division

Co-Investigator(s): *(Co-PI(s)/non students)*

Name: David Comas   Tel No: _____   E-Mail Address: david.comas@upf.edu   ☐

Name: _____   Tel No: _____   E-Mail Address: _____   ☐

Student Investigator(s):

Name: Corrie Avila   Tel No: 508-863-0077   E-Mail Address: cavila@wpi.edu   ☐

Name: Maarja-Liisa Sokk   Tel No: 973-487-0192   E-Mail Address: mlsokk@wpi.edu   ☐

Check if: ☒ **Undergraduate project** *(MQP, IQP, Suff., other)* _____

☐ **Graduate project** *(M.S. Ph.D., other)* _____

Has an IRB ever suspended or terminated a study of any investigator listed above?
No ☒   Yes ☐   *(Attach a summary of the event and resolution.)*

Vulnerable Populations: The proposed research will involve the following (Check all that apply):
pregnant women ☐   human fetuses ☐   neonates ☐   minors/children ☐   prisoners ☐
students ☐   individuals with mental disabilities ☐   individuals with physical disabilities ☐

Collaborating Institutions: *(Please list all collaborating Institutions.)*
National Geographic and IBM - The Genographic Project

Locations of Research: *(If at WPI, please indicate where on campus. If off campus, please give details of locations.)*
Samples collected in small villages in the Veneto hinterland, other potential areas include Paphlagonia (near the Turkish city of Trebizond), Brittany, and NorthWestern Wales, analysis completed at Universitat Pompeu Fabra, Barcelona, Spain

Project Title: Exploring the History of Venice : Relics, Records, and Relations

Funding: *(If the research is funded, please enclose one copy of the research proposal or most recent draft with your application.)*

Funding Agency: National Geographic and IBM - The Genographic Project   WPI Fund: _____

Human Subjects Research: *(All study personnel having direct contact with subjects must take and pass a training course on human subjects research. There is a link to a web-based training course that can be accessed under the Training link on the IRB web site http://www.wpi.edu/Admin/Research/IRB/training.html. The IRB requires a copy of the completion certificate from the course or proof of an equivalent program.)*

Anticipated Dates of Research:

Start Date: 10/24/2010   Completion Date: 12/18/2010

Appendix L: DNA Consent Form

The following are the procedural details of our Genographic Project collaboration.

National Geographic Society in Collaboration with Worcester Polytechnic Institute

Title of Research Project: "*The Genographic Project: Molecular Genetic Analyses of Western/Central European populations"*

National Geographic Principal Investigators: Dr. Jaume Bertranpetit and David Comas, Unitat de Biologia Evolutiva, Universitat Pompeu Fabra. Doctor Aiguader 80 ,08003 Barcelona, Spain.

WPI Venice Project Center Project Title: *"Unearthing the Roots of Venice: From Relics to DNA"*

Overall Project Principal Investigator: Dr. Spencer Wells, Mission Programs, National Geographic Society, 1145 17th Street, N.W., Washington, D.C. 20036, 202-828-5465, SpWells@ngs.org

WPI Principle Investigator: Dr. Fabio Carrera, Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609 USA. Email: carrera@wpi.edu Telephone: +39 041 523-3209

INFORMED CONSENT (Western/Central Europe)

Invitation to Participate: You are invited to participate in a project that studies the various human populations of Western/Central Europe through their inherited genetic properties. This project is sponsored by the National Geographic Society. The project team will talk to you about the study and how it might affect you. They will explain what you will have to do if you decide to participate. You can choose whether or not to participate. If you decide not to participate, no one will take any position of any kind against you.

Only unrelated males can participate, in order to maintain the validity of the study. In addition both maternal and paternal grandparents must have originated from the Veneto region.

Please read this consent form carefully. If you do not understand something, please ask the researcher to explain it. If you prefer, someone will read this to you. If you decide to participate, please sign the last page of this form. We will give you a copy to keep.

Purpose: This project is a major international effort to collect population genetic data from over 100,000 individuals around the world. Our sampling will target native human populations living preferably in their geographic region of origin and who ideally had minimal mixing with the surrounding populations. Our goal is to find out the relationships between genetic, linguistic, cultural and historical data. No data or information on your medical history will be collected from you, and thus the saliva sample collected with the buccal brush will never be used for any medically-related inquiry. To better understand the history of the native people from your region and around the world, the investigators will survey segments of your genetic fingerprint. Changes or variations in your genetic fingerprint may be considered as markers that can be used for comparison among populations. These changes are interesting for the study of origin and behavior of humans because many of them are found to group in descendants of a common ancestor within human populations. Because of the specific way of maternal and paternal inheritance, we can trace these genetic variations through human families from the present to the distant past with a

relatively high degree of accuracy. Therefore, by following the spread of the descendants of a specific ancestry, we can also genetically retrace the patterns of human movement through geographic areas.

The WPI Venice Project Center team is focusing on investigating the genetic heritage of the Venetian people in order to contribute to a better understanding of the origins of Venice. The data collected in the study will be used to help confirm/disprove different theories about the origins of the Venetian people.

Length of the Project: Your participation in the project will require one single visit taking 20 minutes of your time but the entire project will last 5 years. You will be one of many thousands of people in your region who give samples of blood or saliva.

Procedure: If you decide to participate, we will collect a cheek cell sample using a cotton brush. The saliva sample will be used to look for markers in your genetic fingerprint. These markers will be compared to those of ancient native populations. We will also ask questions about your family history.

Storage of Sample: Your saliva sample will be stored indefinitely at the Evolutionary Unit of the Universitat Pompeu Fabra (UPF), where it may be used for further study to better understand human origins and the histories of your people. The saliva sample cannot and will not be used for any medically related study. Furthermore, you may at any point in the future choose to have your sample removed from the project. You can do this by simply contacting our regional center (UPF, Doctor Aiguader 88, Tel. +34 93-3160843) and explaining your desire to have your sample removed. Your sample will then be immediately destroyed at the laboratory and all records and associated data from your sample will be eliminated. Instructions on how to do this are explained below.

Risks: You will spend about 20 minutes on the project. There is almost no risk associated with a cheek swab. Participation is voluntary and we cannot guarantee any results to you.

Benefits: The only benefits to you from participating in this project will be to learn more about your family origins and your relationships to other people around the world.

Costs: You do not have to pay anything to be in the project. You will not be paid for being in this project. If you have to travel to the collecting site, we may pay for your reasonable travel costs. We will not pay for your travel costs, *unless* you arrange for them to be paid for before you provide a blood sample.

Participation: The choice is yours. You may choose either to be in the project, or not to be in the project. Your participation is voluntary. There will be no penalty if you choose not to participate, or if you agree to be in the project but later change your mind. You may withdraw from the project at any time by calling or writing Dr. Jaume Bertranpetit or Dr. David Comas (+34 93-3160843) at the Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Doctor Aiguader 88, 08003 Barcelona, Spain, jaume.bertranpetit@upf.edu or david.comas@upf.edu. You may also withdraw from the project through the website at www.nationalgeographic.com/genographic **.** To withdraw from the project, you must provide your name and assigned code number. Also, the investigators may decide to discontinue your participation in this study without your permission.

Confidentiality: We will protect carefully the information that you tell us about yourself and your family. What we learn from your sample will be described only in a way that does not identify you. You will be given a code number for your sample and instructions how to use the code number to obtain the results of your sample. To protect your privacy, samples will be recorded with a secret code. Your name only will be recorded on the consent form. The secret code assigned to your sample will be kept in a locked file at the site where you participate in the project and carefully protected. No information related to your medical history will be included. Your sample will be stored at the regional site unless you ask to have your sample

destroyed after the project. Your records will be monitored and may be audited without violating confidentiality.

SUBJECT STATEMENT OF CONSENT

I have read and understood the above description of this research study. All of my questions have been answered to my satisfaction. I know that my taking part in this research study is voluntary. I know that I may refuse to take part in or quit this research study at any time. If requested, a copy of this signed consent form will be given to me.

_____   _____

Signature of Subject   Date

_____   _____

Signature of Witness   Date

FOR STUDY USE ONLY

I attest that I have fully and appropriately informed the subject of the nature of the above research study and have offered to answer any question that he/she may have.

_____   _____

Signature of Principal Investigator/Designate   Date

INFORMED CONSENT FORM—ITALIAN

Ricerca internazionale sulle origini degli Antichi Veneti attraverso lo studio del DNA
Una collaborazione tra Worcester Polytechnic Institute (WPI) e National Geographic Society

Nell'ambito del progetto internazionale: "_The Genographic Project: Molecular Genetic Analyses of Western/Central European populations"_

Direttore generale del progetto Genographic: Dr. Spencer Wells, Mission Programs, National Geographic Society, 1145 17th Street, N.W., Washington, D.C. 20036, 202-828-5465, SpWells@ngs.org

Ricercatore Principale di National Geographic: Dr. Jaume Bertranpetit and David Comas, Unitat de Biologia Evolutiva, Universitat Pompeu Fabra. Doctor Aiguader 80 ,08003 Barcelona, Spain.
Titolo della tesi WPI attualmente in corso: *"Unearthing the Roots of Venice: From Relics to DNA"*
Direttore del progetto WPI: Dr. Fabio Carrera, Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609 USA. Email: carrera@wpi.edu Tel: +39 041 523-3209

INFORMAZIONI SULLA PRIVACY PER IL CONSENSO ALLA PARTECIPAZIONE

Invito a Partecipare: La invitiamo a partecipare ad un progetto che studia le diverse popolazioni umane dell'Europa Occidentale/Centrale tramite le loro caratterstiche genetiche. Questo studio è sponsorizzato dalla "National Geographic Society". Il gruppo di studenti che si occupa del progetto Le parlerà dello studio e di come potrebbe toccarLa. Può decidere liberamente se partecipare o meno al progetto e, qualora decisse di non partecipare, ciò non comporterà alcuna conseguenza nei Suoi confronti. Nel caso in cui decidesse di partecipare, Le verrà spiegato dettagliatamente che cosa deve fare.

Solo individui maschi non imparentati tra loro possono partecipare in modo di preservare la validità dello studio. Inoltre è indispensabile che il nonno e la nonna dei partecipanti siano di origini trivenete sia da parte di madre che di padre.

Per favore, legga attentamente questo documento. Se non dovesse comprenderne il contenuto o avesse delle domande al riguardo, chieda liberamente al ricercatore che sarà lieto di offirLe qualsiasi spiegazione. Se preferisce, qualcuno potrà leggerLe questo modulo. Se decisesse di partecipare, per favore firmi l'ultima pagina di questo documento. Le sarà fornita una copia da tenere per sè.

   Descrizione dello Studio: Questo studio fa parte di un grande progetto internazionale che si prefigge di raccogliere dati genetici da oltre100.000 individui in tutto il mondo. I nostri campioni si concentrano su popolazioni indigene che vivono preferibilmente ancora nelle loro terre di origine e che preferibilmente si sono mescolate poco con altre popolazioni locali. Il nostro obiettivo è di scoprire le relazioni tra dati genetici, linguistici, culturali e storici. Non le verrà chiesto alcun dato medico nè informazioni sulla sua salute e quindi il campione genetico che le verrà prelevato non sarà assolutamente mai utilizzato per analisi igienico-sanitarie. I nostri ricercatori analizzeranno segmenti del suo DNA per comprendere meglio la storia delle popolazioni indigene della sua regione e del resto del mondo. Cambiamenti o variazioni nella sua impronta genetica verranno considerati come indicatori da utilizzare per confronti con altre popolazioni. Queste variazioni sono interessanti per lo studio delle origini e dei comportamenti delle stirpi umane dato che permettono di raggruppare popolazioni che discendono da antenati comuni. Grazie ai meccanismi di ereditarietà sia dal lato materno che paterno, possiamo ricostruire con discreta precisione queste variazioni genetiche attraverso stirpi umane dal presente fino alla piu remota antichità. Quindi, seguendo la distribuzione dei discendenti di una specifica stirpe, possiamo ricostruire i percorsi migratori seguiti dalle varie popolazioni attraverso varie aree geografiche.
Lo studio specificamente condotto dal Venice Project Center del Worcester Polytechnic Institute nel contensto del piu ampio progetto *Genographics*, si concentra esclusivamente sul patrimonio genetico dei Veneti in generale e dei Veneziani in particolare, con l'intento di contribuire ad una migliore comprensione delle origini di Venezia. I dati raccolti da questo studio serviranno a confermare o smentire diverse teorie sulle origini geografiche dei Veneti, sulla loro diffusione

nel continente europeo e soprattutto sul loro insediamento nelle lagune dove oggi sorge Venezia.

Durata dello Studio: La partecipazione allo studio comporterà un solo incontro con un addetto della durata di circa 20 minuti, mentre l'intero progetto avrà una durata totale di 5 anni. Lei sarà uno delle migliaia di persone della sua stessa regione a donare un campione di sangue o saliva.

Procedura: Qualora decidesse di partecipare, Le chiederemo di raccogliere un campione di saliva prelevandolo da sè dall'interno della guancia utilizzando uno strumento simile ad un piccolo spazzolino da denti con l'estremità tipo cottonfioc. Il campione verrà poi utilizzato per identificare alcuni marcatori specifici nella sua impronta genetica. Gli indicatori genetici così raccolti saranno comparati con quelli delle popolazioni native della regione. Le verrano inoltre richieste alcune brevi informazioni riguardanti la storia della Sua famiglia.

Stoccaggio del campione: Il Suo campione di saliva sara preservato indefinitamente dall'Unità Evoluzionaria dell'Università Pompeu Fabra (UPF) di Barcellona, dove potrebbe essere utilizzato per ulteriori studi per la comprensione delle origini storiche delle popolazioni mondiali. Ciononostante, sarà Sua facoltà di decidere di rimuovere il Suo campione dal progetto in qualsiasi momento futuro. Per far questo, basterà contattare il nostro centro regionale (UPF, Doctor Aiguader 88, Tel. +34 93-3160843) esprimendo il Suo desiderio di rimuovere il Suo campione dallo studio. Il Suo campione verrà allora immediatamente distrutto nel laboratorio e tutti i dati ad esso associati verranno anch'essi eliminati. Istruzioni dettagliate su questa procedura sono illustrate in calce.

Rischi, Inconvenienti e Disagi: La Sua partecipazione comporterà un impegno di tempo totale di circa 20 minuti. Al momento non si conoscono rischi per la salute nell'eseguire un prelievo di un campione di DNA. La partecipazione è volontaria e non ci è possibile garantirLe la buona riuscita del test genetico.

Benefici: L'unico reale beneficio derivante dalla partecipazione a questo studio sarà quello di avere maggiori informazioni riguardanti le origini della Sua famiglia e le relazioni tra questa e altre persone nel mondo.

Costi: Nessun costo è dovuto per la partecipazione al progetto. Non c'è alcuna ricompensa per la partecipazione.

Partecipazione: La Partecipazione a questo studio è volontaria. Puoi scegliere di non partecipare. Hai il diritto di ritirarti in qualsiasi momento e decidere se il tuo campione biologico, le informazioni genetiche o la genealogia già raccolti restino a far parte di questo studio o vengano distrutti.
Puoi decidere di ritirarti dal progetto in qualsiasi momento chiamando o scrivendo a Dr. Jaume Bertranpetit or Dr. David Comas (+34 93-3160843)alla Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Doctor Aiguader 88, 08003 di Barcelona, Spagna, jaume.bertranpetit@upf.edu oppure a david.comas@upf.edu. Può ritirarsi dal progetto anche attravero il sito web www.nationalgeographic.com/genographic. Qualora desiderasse ritirarsi dal progetto, dovrà comunicarci il suo nome e codice segreto identificativo assegnatoLe. Inoltre, i ricercatori del progetto potrebbero decidere di escludere il Suo campione dallo studio senza la Sua esplicita autorizzazione.

Riservatezza: La Sua identità, unica e personale, è considerata strettamente riservata e privata. La Sua identità, unica e personale, non sarà rivelata in alcuna pubblicazione generica o scientifica di dati. Campioni e documenti contenenti le Sue informazioni saranno tenuti in un luogo sicuro (vedere il paragraforelativo allo stoccagio del campione). Le uniche persone ad avere accesso ai codici e alle informazioni genealogiche saranno l'investigatore principale e le altre persone specificamente autorizzate dall'investigatore principale. La Sua identità unica e l'identità dei Suoi antenati recenti non sono collegate direttamente alle informazioni contenute nel database pubblico.

Le sarà assegnato un codice numerico e le istruzioni di utilizzo del codice stesso per poter accedere ai risultati relativi al Suo campione. a protezione della Sua privacy, il campione sarà quindi schedato con un codice segreto. Il Suo nome sarà registrato solo all'interno del presente Modulo di Consenso. Il codice segreto assegnato al Suo campione verrà attentamente custodito sotto chiave, all'interno di uno schedario appositamente protetto da intrusioni, presso la sede centrale del progetto. Nessuna informazione relativamente alla Sua situazione medica sarà inclusa. Il suo campione sarà custodito all'interno del sito locale relativo alla sua regione a meno che Lei non richieda esplicitamente la sua distruzione al termine dello studio. I suoi dati saranno sorvegliati ed occasionalmente visionati da addetti ai lavori senza che questo comporti la violazione della Sua privacy.

CONSENSO

Se è d'accordo a partecipare a questo studio, riceverà una copia firmata e datata di questo Modulo di Consenso per Suo uso personale.

Ho letto le informazioni su questo Modulo di Consenso. Ho ricevuto tutte le risposte alle mie domande in relazione allo studio e alla mia partecipazione. Sono consapevole che la mia partecipazione in questo studio di ricerca è frutto della mia volontà. Sono consapevole di avere il diritto di rifiutare di partecipare e anche di ritirarmi da questo studio in qualsiasi momento. Su richiesta, mi sarà fornita una copia di questo consenso firmato.

_____   _____
Firma del Soggetto   Data

_____   _____
Firma del Testimone   Data

AREA RISERVATA AI CONDUTTORI DELLO STUDIO

Attesto di avere completamente informato il soggetto circa la natura di questa ricerca e mi sono reso disponibile a rispondere a qualunque domanda che il soggetto avesse al riguardo.

_____   _____
Firma del Capo Ricercatore o del suo Delegato   Data

Appendix M: Project Information Card

WPI - Venice Project Center
National Geographic

Riconstruzione delle Origini dei Veneziani attraverso
L'Analisi del DNA

NATIONAL GEOGRAPHIC

Per Ulteriori Informazioni
Contattare Venice Project Center:

Tel: +39 041 523-3209
Email: carrera@wpi.edu
http://venice2point0.org

Per accedere ai
risultati delle sue
analisi genetiche
vedere istruzioni a
tergo

I risultati sono protetti da un codice segreto personale per garantire la massima riservatezza

Per Accedere ai Risultati delle sue Analisi

1. Accedere al sito web:
   https://genographic.nationalgeographic.com/genographic/lan/en/journey.html
2. Dopo la scritta "Enter Your Genographic ID", inserire il codice sectreto personale:
3. Poi premere "Invio"
4. I suoi risultati appariranno sullo schermo
5. Qualora I risultati non siano ancora pronti riprovare alcuni giorni dopo. I risultati saranno consultabili dopo circa due mesi dalla data del prelievo

Appendix N: Funding Letter for Genealogy

Exploring the History of Venice: Relics, Records, and Relations

Dates
- Project Center Dates
  - October 24 - December 18, 2010
- Genographic Project
  - 5 year research partnership by National Geographic and IBM
  - ends in 2011

Participants
- Corrie Avila
- Elizabeth DeZulueta
- Sofia Mukhanov
- Maarja-Liisa Sokk

Faculty Sponsor
- Fabio Carrera
- James Cocola

Introduction

The Interdisciplinary Qualifying Project (IQP) is a key piece of the curriculum at Worcester Polytechnic Institute. Each project is assigned to a team of four students, with initial research completed on campus and the remainder of the project carried out off-site at a location relevant to the research being done. Our team has been accepted into the Venice Project Center Program where we will be researching and exploring the history of Venice and the origins of the ancient Veneti. Our project is multifaceted and has three main components: written historical documents, archaeological data, and genetic genealogy. This letter focuses on the genealogical aspect of our project and its respective components.

Project Description:

Our project is aimed at discovering the origins of the ancient Veneti through genetic genealogy. National Geographic and IBM are sponsoring the Genographic Project, an effort to determine the ancestry and origins of individuals and global populations through the analysis of DNA samples. Although there is evidence of Veneti influence throughout Europe, their actual origin is still undetermined. Our team will be investigating the three major regions that are theorized to be the origin of the ancient Veneti: Lusatia (modern-day Poland and Germany), Brittany and Wales (north-western France and United Kingdom), and Paphlagonia (modern-day northern Turkey). By collaborating with the National Genographic Project, we will aim to collect DNA samples from people native to each of these regions as well as small villages in the Veneto. The collected DNA samples will be sent to the Genographic Project laboratory in Barcelona, Spain in order to be analyzed and compared to find genetic correlations. Using the results of the analysis, we hope to gain insight into the ancestry of the people of Venice.

DNA Collection Procedures

Our team will focus on collecting DNA samples from native residents of the previously mentioned regions in order to investigate their genetic histories to help confirm or disprove different theories about Veneti origins. We will aim to collect 125 samples of DNA from locals whose families have preferably resided in the local area for an extended* number of generations. According to the Genographic Project's guidelines for DNA collection, we will be limiting participation to unrelated males who:

- are over the age of 18,
- are of "given" ethnicity,
- were born in the "given" region,
- have both biological parents of the "given" ethnicity,
- have both maternal grandparents of "given" ethnicity,
- have both paternal grandparents of "given" ethnicity,

where "given" denotes one of the theorized areas of testing--Paphlagonia, Brittany, Wales, Lusatia, or Veneto.

To maximize the accuracy and usefulness of the DNA samples, participants will be required to fill out forms regarding their identification and familial history. Personal health information and medical history will *not* be collected from individuals, and the DNA samples will never be used for any medical-related purposes. Individual genetic analyses will *not* be publicly disclosed, unless inquired about by the participant. We will collect participants' DNA by administering a *Whatman Sterile Omni Swab* cheek swab sampling method, as described on the following page.

Whatman Sterile Omni Swab
Instructions For Use

The Whatman Sterile **Omni Swab** is designed for collection of buccal cell samples for DNA testing. To obtain a buccal cell sample, the tip end of the **Omni Swab** is rubbed against the inside of the cheek. Unlike cotton-tipped swabs, the **Omni Swab** is made of absorbent material specifically designed for the collection of buccal cells. The single use swab has a unique removable collection tip to assist in processing the sample. While collecting a buccal sample is quick, easy and painless, it needs to be done correctly.

Suggested Directions for Use
The following directions are provided to illustrate an acceptable collection protocol. Each facility should establish a sample collection protocol that meets the objectives of collection and testing facilities.

Sample collection:
The person providing the buccal cell samples should not eat or drink immediately prior to giving the sample. If food or drink has been recently taken, it is suggested that the mouth be rinsed with water prior to sampling. The person taking the samples should wear biohazard-barrier gloves and avoid contact with the *Omni Swab* collection tip.

1. Open the *Omni Swab* packaging at the handle end and carefully remove swab. Do not touch the collection tip of the swab.

2. Holding the handle end of the *Omni Swab*, scrape the collection tip firmly against the inside of the cheek 5-6 times (about 10 seconds) being careful not to depress the plunger that ejects the tip.

3. After taking the sample, eject the tip by firmly pressing the plunger at the end of the handle into a labeled 2 mL microcentrifuge tube.

4. If desired, repeat sampling procedure with a second swab using the other cheek. Eject tip into a different labeled 2 mL microcentrifuge tube.

Storage:
After sample collection, tip can be kept at room temperature when processed immediately. If storage is necessary, freeze tips at -20°C.

DNA Extraction:
DNA can be extracted from the tip using standard laboratory procedures. DNA yields of 500-2,000 ng DNA are typical. Actual DNA yield will vary depending on the original DNA concentration, collection efficiency and extraction procedure.

Caution: Read instructions prior to use.
Single use only. Do not use if package has been previously opened or damaged. Federal law (USA) restricts this device to sale by or on the order of a physician.

To obtain Whatman® Sterile *Omni Swabs*, Catalog #WB10-0004. Contact Whatman® toll free at 1-866-787-3362.

Whatman®
www.whatman.com

Methodology and Statements of Need

Both financial and non-financial aid is needed in order for our team to successfully complete this project. Since our work entails collecting, storing, and analyzing DNA samples from various relevant cities throughout Europe, is it essential to get both funding and collaborator support for our efforts.

To devise a plan of DNA collection in Paphlagonia, Wales, Brittany, and Veneto, we are currently collaborating with several local contacts who are providing us with information about the regions and their useful resources. However, our team could use more support to better identify local organizations or premises within the relevant cities to find participants for DNA collection. As previously mentioned, we are targeting elderly males whose families are native to the regions. Finding a strategic method or location to attract participants is a crucial aspect of our work and our team will need as much help as possible since our team is unfamiliar with the mentioned regions.

In order to successfully administer DNA collection throughout the relevant regions in Europe, our team will need financial aid for travel and its associated expenses. Although the ultimate goal is to travel to all of the mentioned regions and collect DNA in cities throughout each location, the limited amount of time we spend in Europe restricts our team to travel to only two or three of the regions. Funding is necessary for transportation to, from, and within each region in order to successfully complete our work. More specifically, money is needed for round trip airplane tickets to a relevant country, local public transportation by bus, cab, or shuttle, and various other transportation expenses that will accumulate during DNA collection. Besides the costs of transportation, we will be paying for living accommodations, food, and various basic everyday necessities.

Based on advice given from one of the individuals we are collaborating with to plan DNA collections in Paphlagonia (Turkey), local participation may be encouraged by a monetary reward system. Because our goal to collect 125 samples from each relevant region may be difficult to attain, a marketing plan to attract participants will be more effective if combined with a small monetary compensation or its equivalent. Examples that were mentioned by a collaborator include snacks, beverages, small souvenis, etc.

In the case(s) that our team will not be able to travel to the relevant regions to administer DNA collection, corresponding work will be delegated to local collaborators who are willing to contribute their efforts to our project. Our team will need help to both locate specific sites for collection, find valid participants, and then administer DNA collection following the provided procedure.

In both cases, after DNA collection, the samples will need to be stored in appropriate conditions and then shipped to the Genographic Project Center laboratory in Barcelona, Spain. Financial support for both shipping and storage costs will be necessary for the preservation of the samples.

Contact Information
The best way to contact our project group would be by email at ve10-orig@wpi.edu.