Geospatial Innovation in the Digital Humanities:

Implementation and Evaluation of Deep Mapping in the Lake District

Alexander Reinhold M.Sc. (Merit)

School of Computing and Communications

History

Lancaster University

A thesis submitted for the degree of

Doctor of Philosophy

September 2020

Abstract

This thesis explores the concept of Deep Mapping and how it can be implemented and evaluated using historic texts from the Lake District region. Deep Mapping is a new way to approach, understand and analyze the relationship between geography, history and literature through a variety of media. Deep Mapping has been discussed in the literature, but few practical implementations of the concept have been created. This thesis works to bridge the gap between the theoretical concept of Deep Mapping and its real-world application and use. The primary data used for this thesis is the Corpus of Lake District Writing, a collection of texts collected by researchers. This corpus contains 80 manually digitized texts about the Lake District region from 1622 to 1900. The corpus is made up of several different genres, including travel journals, essays, novels, and poetry. These texts, along with other multimedia, were brought together with interdisciplinary methods drawn from the fields of history, literature, corpus linguistics, spatial and digital humanities and computer science to create Deep Mapping prototypes.

A series of case studies, discussed in chapters 3-7, were undertaken to explore Deep Mapping. Each case study focuses on developing a different Deep Map prototype addressing a new problem and set of objectives. Each prototype was then evaluated based on three sets of criteria: addressing the problem and objectives (design), this thesis's research questions (research), and classifications of digital and spatial humanities and Deep Mapping (data). The last case study, chapter 7, focuses on processing data not related to the Lake District in order to explore if the methods used to develop the prototypes in chapters 3-6 are repeatable with other data sources. The final chapters discuss the implementation and evaluation of Deep Mapping. This thesis produced four Deep Mapping prototypes that can be used as exemplars for future research. The evaluation of these prototypes led to the development of the Evaluation Rubric, that can be used in the evaluation of applications to help determine their fit as a Deep Map and to assist in the comparison in Deep Mapping applications developed in contrasting forms and based on differing content.

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university.

Alexander Reinhold

Acknowledgements

This research received funding from the Leverhulme Trust (RPG-2015-230), as part of the 'Geospatial Innovation in the Digital Humanities' project.

I would first like to thank my supervisors, Professors Ian Gregory and Paul Rayson, whose expertise was invaluable. Their support and supervision gave me the resources and direction needed to pursue my research. I appreciate them for the opportunities they provided during my research; traveling for conferences and assisting in teaching were invaluable experiences. Thank you for your continuous support.

I would like to acknowledge my colleagues from the Geospatial Innovations project for their wonderful collaboration. Their support and feedback was instrumental in defining the path of my research. I would particularly like to single out Chris Donaldson for his assistance in conceptualizing the first case study of this thesis and his collaboration on my first published article on that work.

I would also like to thank Anne Knowles for including me in the Holocaust Testimonies processing project. My contribution to the project helped inform my research and expand it beyond the regional focus of the Geospatial Innovations project.

In addition, I would like to thank my parents and brother for their eternal support. They are always there for me. Finally, I could not have completed this thesis without the support of my friends, who provided happy distractions to rest my mind outside of my research.

Publications

The following have been published as part of the research presented in this thesis. Where appropriate, portions of this thesis are based on my contributions to these publications without citation. Where research and text should be credited to a co-author, rather than myself, the work has been cited accordingly. Work from 'A deeply annotated testbed for geographical text analysis' appears in chapter 2.2.2. Work from 'Exploring Deep Mapping Concepts: Crosthwaite's Map and West's picturesque Stations' appears in chapter 3. Work from 'Deep Mapping Tarn Hows: Automated generation of 3D Historic Landscapes' appears in chapter 6.

Rayson, P. E., **Reinhold, A.**, Butler, J., Donaldson, C. E., Gregory, I. N. & Taylor, J. E., 7/11/2017, A deeply annotated testbed for geographical text analysis: The Corpus of Lake District Writing. *GeoHumanities'17 Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities.* New York: Association for Computing Machinery (ACM), p. 9-15 7p.

Reinhold, A., Donaldson, C. E., Gregory, I. N. & Rayson, P. E., 14/09/2017, Exploring Deep Mapping Concepts: Crosthwaite's Map and West's Picturesque Stations. *Proceedings of Workshops and Posters at the 13th International Conference on Spatial Information Theory* (COSIT 2017). Fogliaroni, P., Ballatore, A. & Clementini, E. (eds.). Cham: Springer, p. 265-273 9 p. (Lecture Notes in Geoinformation and Cartography).

Reinhold, A., Gregory, I. N. & Rayson, P. E., 2018, Deep Mapping Tarn Hows: Automated Generation of 3D Historic Landscapes. *Eurographics Workshop on Graphics and Cultural Heritage.* Sablatnig, R. & Wimmer, M. (eds.). The Eurographics Association

Contents

List of Figures

List of Tables

List of Abbreviations

API	Application Programming Interface
CLDW	Corpus of Lake District Writing
GIS	Geographic Information Systems
GTA	Geographic Text Analysis
GUI	Graphical User Interface
HGIS	Historical Geographic Information Systems
HTML	Hypertext Mark-up Language
IDE	Interactive Development Environment
JSON	JavaScript Object Notation
KML	Keyhole Markup Language
KMZ	Zipped KML file
ML	Machine Learning
NER	Named Entity Recognition
NLP	Natural Language Processing
POS	Part of Speech
RtD	Research Through Design
SQL	Structured Query Language
UCREL	University Centre for Computer Corpus Research on Language
XML	Extensible Markup Language

1 Introduction

This PhD is part of the Geospatial Innovation in the Digital Humanities: A Deep Map of the Lake District project, funded by the Leverhulme Trust. The goal of this project is to develop new insight into the literary and cultural geographies of the English Lake District region, through the use of interdisciplinary research in spatial humanities. This project focuses on bringing together methods from computer science, geographic information science, literary studies, and history to change the way geography is employed in order to identify space and place.

This thesis evaluates the practical implementation of Deep Mapping. Deep Mapping is a relatively new approach to visualizing, understanding and analyzing the relationships between geography, history and literature, using a variety of multimedia to explore the multiple layers of meaning of a place. A Deep Map is an interactive tool and spatial representation of the "the multiple histories of place, the cross-sectional stories of natural and human history." While traditional maps are generally static and focus on the depiction of space at a particular time, a Deep Map is meant to be a depiction of a place over a period of time.

This research uses the English Lake District in the United Kingdom as the subject to explore Deep Mapping methods. The Lake District is one of Britain's most significant cultural landscapes, with a rich literature and cultural history. The core source of data for this research is the Corpus of Lake District writing, a collection of texts originally composed between 1622 and 1900, representing a range of genres and authors, with a focus on the Lake District.

Four prototypes were developed for this research to evaluate various aspects of Deep Mapping. Each Deep Map prototype, described in chapters three through six, is designed to demonstrate Deep Mapping concepts in varying ways to target different user groups. The development of these prototype Deep Maps allows a range of user groups to engage with the data, in order to gain new understandings of space and place in the Lake District.

1.1 Significance of the Problem

This research is significant because while there is a large range of data available, it is not always meaningful in its raw form. Deep Mapping is a solution to this problem, allowing different user groups to engage with the data, to better understand and analyze it, while maintaining the historic context of the material.

While there is a strong theoretical base for Deep Mapping, few practical implementations have been developed. This study looks at the theoretical elements of Deep Maps and explores what is achievable in a practical application. Deep Maps have the potential to enhance the research process and provide engagement with data that may lack meaning to different user groups in its raw form. A Deep Map provides a platform to explore the spatial relationships between potentially disparate data. Current implementations of Deep Maps tend to be relatively simple, concentrating on data linked to markers on a map, as seen in the Cork and ICA maps discussed in chapter 2.4.2. While this can fit the definition of a Deep Map, basic implementations do not explore the many possible forms a Deep Map could take or allow users to explore the complex interconnectedness of the data. The prototypes built as part of this thesis's research attempt to move beyond the map and marker model and create a more complex, yet still usable and understandable, interaction with the data in a spatial environment.

1.2 Analyze the Theoretical Basis for the Study

The research presented in this thesis aims to develop and evaluate Deep Map prototypes. This research uses the research through design methodology to:

- Identify problems of interest
- Establish objectives for a solution
- Design and development of prototypes to address the problem and create a solution based on the established objectives
- Demonstrate the prototypes to allow iterative feedback
- Evaluation based on objectives of solution and problem constraints, and larger Deep Mapping evaluation.
- Communication of results.

Evaluation of each case study's prototype developed for this research is based on three categories.

Design Evaluation - The first set of criteria for evaluation is drawn from each case study's established problem requirements and objective of the designed solution. Each prototype has a defined problem it is seeking to address, with a set of objectives to be included in the designed solution. The problem and objective will be the first measure of how well the resulting solution meets its intended goals.

Research Evaluation - The second set of criteria are focused on each case study ability to address the research questions of this study. Each case study, in addition to addressing its own defined problem, is targets at answering at least one of the research questions of this thesis.

Data Evaluation - The third set of criteria are focused on determining how well each case study fits within the definition of Deep Mapping. This set is a combination of qualitative and quantitative questions, that uses various user groups responses to form a consensus on what elements within each developed prototype are meaningful to the concept of Deep Mapping. The goal of this set of criteria is to establish a viable method of evaluating future Deep Map applications. Data for this step of evaluation was collected from an Evaluation Survey, described in section1.6.1, and in person demonstrations at public events, university seminars, and conference presentations. This data was used to complete an Evaluation Rubric for each case study. The rubric is made up of four subcategories, Digital Humanities, Spatial Humanities, Deep Mapping, and Research Questions.

The Digital Humanities subcategory contains criteria drawn from the values and methods central to the Digital Humanities, as described in section 2.1.2. The criteria for the Digital Humanities subcategory look at how iteration and experimentation are used in the development process (Iterative Process and Experimental); how collaboration helps to inform iteration and create more inclusive development (Collaborative); and how accessible the results of research and development are to the various user groups (Open and Accessible), either through design of the applications or dissemination of research through publications and presentations (Multimodal). These criteria were chosen because they are important to understanding the research and development of applications within the field of Digital Humanities and they overlap with common practices in computer science for development of digital applications, providing common ground to comparing Deep Mapping applications across the digital space.

The Spatial Humanities subcategory contains criteria drawn from concepts at the core of the field, as described in section 2.1.1. The criteria for this subcategory look at how the prototypes address the spatial and temporal elements of data (Relationship between space and time), and how the concept of 'Place' can be constructed using multiple layers of data to create a sense of connection and meaning for users engaging with the material (Place built through multilayered data). These criteria overlap with definitions in Deep Mapping and allow for the comparison of Deep Mapping applications to other spatial systems like GIS that support Spatial Humanities research focused on the concepts of space, place and time.

The Deep Mapping subcategory includes criteria that address the various definitions of Deep Mapping that have been proposed in the literature, as described in section 2.4. The criteria for this subcategory look at how well an application meets the definitions of Deep Mapping. The criteria look at whether an application allows for exploration using spatial navigation as a primary form of interacting with the application and the data it contains (Spatial Exploration as a primary form of interaction). The temporal component of data, and how the element of time is used in the discovery of data in the application is also assessed as it relates to navigation and representation of data in the application (Temporal context as a primary form of discovery). The criteria look at whether the data used within the application is derived from valid sources, those that can be academically and scientifically supported (Data from verifiable source for legitimacy). An example of a verifiable source would be the CLDW, which was assembled by literary researchers from published historic texts, where an unverified source would be something like public comments (ex Twitter) without any curation to determine relevance and authenticity. The criterion for this subcategory also determines if the data used fills the extents, the physical bounds, of the application (Layers fill targeted area). This is important because the creation of a Deep Map that had the entire world as its target area, but only included data from a single country, would look empty and would have a lack of material to engage users across the targeted area. The final component that is assessed by this subcategory is the accessibility of evaluated applications to various audiences, looking at the ease of use and the potential barriers to entry for using the application, since these affect how inclusive a Deep Map application can be (Accessible to broad audience).

The last subcategory of the evaluation rubric is based on the research questions of this thesis, described below in section 1.5, and some broader research concepts important to the production of academic tools. This subcategory was included in order to provide comparable results across the case studies. While the Research Evaluation category, described above, focuses on a qualitative discussion of each case studies attributions to this thesis's research, this subcategory provides a more quantifiable view. The criteria for this subcategory, based on RQ1, looks at the application's impact on the research process and its potential contribution to research (Inform Research Process). The criteria also consider concepts based on RQ2, looking at how users use the applications and how effective the application is at disseminating data and involving uses in the exploration of the material (Accessible and understandable, Engaging). Aside from the research questions, the criteria of this category also look at the repeatability and expandability of the applications (Repeatable, Expandable). This is important to understanding how the methods and practices used to develop the prototypes can be used in the future. If those methods and practices are not repeatable or don't allow for expansion, in scope or data, the impact of this research would be limited to the contributions it makes to the concept of Deep Mapping, and not on the practical implementation of Deep Mapping.

The three categories Design, Research, and Data were chosen in order to approach the evaluation of the prototypes from different sides. Deep Mapping definitions allow for a variety of possible forms a Deep Map can take, so a single set of criteria for evaluation would either be too vague to offer insight or too strict, excluding more creative interpretations of the existing definitions. Using three sets of evaluation criteria allows each prototype to be evaluated for their effectiveness as a Deep Map and allow comparison between prototypes, while still allowing for flexibility in the design of the applications.

1.3 Pertinent Literature

Deep Mapping is described as a platform, process, and product for bringing together data in a shared environment, with a focus on space and time (Bodenhamer 2015). Deep Maps use spatial exploration (Ridge 2013) and multilayered data (Oxx 2013) to create a sense of place as discussed in chapter 2.1. At the intersection of Spatial and Digital Humanities, Deep Mapping is a multidisciplinary approach to the spatial organization, analysis and dissemination of data.

The word Deep in Deep Mapping can be seen in two ways according to the various definitions provided in academic literature. Deep can refer to the depth of content in an application, concentrating on multilayered design and exploration of data through connections created in the application. Deep can also refer to depth of meaning, concentrating on the concepts of Space vs Place (chapter 2.1). A space is a location without attached meaning, so a Deep Map, a map of place would aim to convey meaning. This meaning can be conveyed through personal accounts, visual records, or any media that creates a meaningful connection between the user and the location.

While the word Map(ping) implies that a Deep Mapping application should be based on a cartographic representation of Earth, a map, the definitions of Deep Mapping allow for a looser interpretation. Deep Mapping, as described in academic literature, must allow for spatial exploration but not explicitly be a geographic map (chapter 2.4). The looseness of the definition allows for the potential for Deep Maps to be based on more abstract representations of space like the Prezi example discussed in chapter 2.4.2, or the 3D Tarn Hows case study discussed in chapter 6.

1.4 Problem Statement

The purpose of this research is to evaluate the practical implementation of Deep Mapping. This research looks to understand how Deep Mapping can be constructed with an underlying historic dataset and identify what visualization elements allow for dissemination and analysis of that data.

The core of the PhD is to develop a series of prototype Deep Maps, envisaged by the Geospatial Innovations team as intuitive, open access web tools, that allow a range of user-groups to gain new understandings of the importance of space and place to Lake District heritage. The core source will be the Corpus of Lake District writing – over 1.5 million words from texts written between 1622-1900 from which place names have been extracted and geo-located. These texts have a strong concentration on a writer's experiences of real-world locations in the Lake District. Deep Mapping offers a new way to approach, understand and analyze the relationship between geography, history, and literature through a variety of media that allows for full exploration of multiple layers of meaning in relation to the object of study.

1.5 Research Questions and Hypotheses

This thesis has three main research questions, designed to explore the practical implementation and evaluation of Deep Mapping.

> **RQ1** - How can Deep Mapping, in combination with spatial analysis, natural language processing, and corpus linguistic techniques, be used to inform the research process?
>
> **RQ2** - What elements are required to best present historic data to various user groups in a way that makes it accessible and understandable? What visualization techniques best engage these different user groups?
>
> **RQ3** – How can Deep Mapping applications be evaluated given the vast array of forms a map might take under the loose definitions of Deep Mapping?

The user groups targeted by this study are:

> **UG1** - Researchers and scholars, specifically those in literature and history, or those with an interest in the Lake District.
>
> **UG2** - Site authorities and Heritage professionals who have a need for a practical tool for the dissemination of data or conservation of historic information.

UG3 - The general public, including tourists and the local community with an interest in the Lake District.

The first research question is designed to help better understand the use of Deep Mapping in the academic process. There are many methods and practices used across the multidisciplinary fields of Spatial and Digital Humanities that are often used in isolation. RQ1 looks at how these methods and practices can be brought together to enhance existing research and to provide new insights that might create new research opportunities. As a tool that brings together technologies and methods from different fields, in theory Deep Maps can provide a platform for informing the research process, but practical implementations of Deep Map applications have been rare and often focus more on the presentation of data rather than creation of a tool to inform research. The prototypes developed for this thesis will help answer RQ1 and provide insight into how Deep Maps can be used as tools for academic research. This research question focuses heavily on the first user group (UG1) targeted by this research but could have value to UG2 as well given that there can be significant crossover between research practices and activities like site conservation.

The second research question is designed to explore the common elements that various forms of Deep Maps might share. There are many forms a Deep Map could take and many ways that data can be presented to users. RQ2 aims to determine what elements are effective in providing data to users in a way that allows them to understand it and actively engage and interact with the material. Each of the prototype applications developed through this research attempt to present the same or similar data to users in different ways, providing different visualizations of the data and different methods of interaction. By answering RQ2, a set of common elements that are effective parts of a Deep Map will be identified. While the flexibility allowed by Deep Mapping definitions means this set of elements might not be shared by all Deep Maps, these common elements will provide a set of building blocks that future iterations of Deep Maps can choose to use to meet the basic needs of a Deep Mapping application. This research question applies to all of the user groups targeted by this research, as providing understanding and engagement is important for all these groups. The prototypes developed for this research general try to be understandable and engaging to all user groups at the same time. Some, specifically the Analytics Deep Map, described in chapter 4, have a greater focus on engaging academic audiences and some of the more complex aspects of the application are not as accessible to other audiences.

RQ3 is key to understanding how Deep Mapping applications can be assessed in a shared space. Section 1.2 introduces the three evaluation categories used to assess the Deep Mapping applications developed for this research. These categories provide insight into what a Deep Map is and how different applications can be compared to one another based on their implementation as a Deep Map, even when the form and content do not have significant overlap.

The research questions (RQs) and user groups (UGs) stated here are referred to through this thesis in order to discuss how they relate to the case studies and how the prototypes contribute to addressing each research question. Each case study also defines case study problems (CSPs) and case study objectives (CSOs) which are unique to each case study and are used to evaluate the design elements of the prototypes created.

1.6 Evaluation of Deep Mapping Applications

This section discusses the evaluation of the prototypes developed for this research and the potential of evaluating other Deep Mapping applications, using these prototypes as exemplars. While maps and mapping tools have been widely available, there is no generally accepted benchmark for comparison and evaluation (Bellahsene 2011). Evaluation of traditional maps and GIS outputs generally focus on composition and design of the map, like legends and clarity, or on the intangible benefits they create, like improved communication or improved access to data. These intangible benefits are evaluated by looking at outcomes that are expected of the map's use. For example, a map that is used in reports or presentations meets the expected outcome of a map seeking to improve communication (Kurwakumire 2014). This type of evaluation is useful in instances where maps are produced as deliverables for projects and disseminated to concerned parties, like in corporations or government agencies. Evaluation of Deep Mapping proves to be a more difficult task as the deliverable is a dynamic and interactive map. Deep Mapping is a broad concept that can be implemented in many different forms that make comparison and evaluation problematic. This thesis develops an evaluation methodology and applies this to evaluate the prototypes developed for the case studies by looking at three aspects of each implementation:

Design – How well does the design of the map fit the problems and objective of the project?

Research – How well does the map address the research questions stated in section 1.5 of this thesis?

The research questions of this thesis are:

RQ1 - How can Deep Mapping, in combination with spatial analysis, natural language processing, and corpus linguistic techniques, be used to inform the research process?

RQ2 - What elements are required to best present historic data to various user groups in a way that makes it accessible and understandable? What visualization techniques best engage these different user groups?

RQ3 – How can Deep Mapping applications be evaluated given the vast array of forms a map might take under the loose definitions of Deep Mapping?

Data – How well does the map fit the definitions of Deep Mapping? What user feedback supports that? And how well does the map fit within Spatial and Digital Humanities methods and practices?

1.6.1 Design Evaluation

Each case study has a discussion on how that individual case study addresses the three categories, which can be found in sections 3.6, 4.5, 5.5, and 6.6. The prototypes for each case study were designed to meet a set of stated objectives in order to address the problem that motivated the case study. All the prototypes fulfilled their stated objectives and so the design category of evaluation was successful for all case studies. While the design category of evaluation was unlikely to produce negative results given that each prototype was designed to address criteria unique to that case study, the design category of evaluation still has value because it helps to demonstrate the exploratory, experimental, and creative aspects of building Deep Mapping applications.

1.6.2 Research and Data Evaluation

This section looks at how the Deep Map prototypes were evaluated using user feedback from an evaluation survey and live demonstrations. This section also looks at how the research questions of this thesis informed the evaluation, by including criteria based on those questions in the evaluation rubric. A discussion of how the research questions were addressed by the research undertaken for this thesis can be found in section 8.2.

Each case study chapter individually discusses how well the designed prototype addresses the research questions from this thesis. To evaluate and compare all of the prototypes based on the research questions, a section of the evaluation rubric, discussed below, contains criteria that addresses the questions as well as break some aspects out by user group.

1.6.3 Evaluation Survey

The evaluation rubric found in the Data Evaluation section of each case study, was completed for each case study using feedback from in person demonstrations, seminar and conference presentations, and an evaluation survey. The evaluation survey was conducted after the completion of all four prototypes. To conduct the survey a custom survey tool[1] was built so that users could answer questions while using the prototypes. The survey allowed users to select which prototypes they were interested in viewing, and while they explored each application questions would be asked based on the activities the users were performing. This method of performing the survey enabled users to freely explore the prototypes, without direction, a key element to Deep Mapping, and still provide the functionality to gather feedback from the users. The most significant limitation to this method of survey was that not all users would receive all the questions if they did not explore all aspects of a map.

TABLE 1 SELECTION OF SURVEY QUESTIONS

Survey Questions
Rate the look of this map
How useful is multimedia in engaging with this location?
Any General Comments?
How well does changing location help understand the text
How well does flipping the pages engage you with this experience?
How useful is the panorama image?
Deep Mapping is a concept based on the spatial representation of the various meanings and experiences of a place over a period of time. Would you consider this a Deep Map?
How useful is this search tool?
How useful is this advanced search tool?
How useful is this compare search tool?
How useful is this advanced compare search tool?
How useful is this location search tool?
How useful is the Historic map in context?
How useful is a timeline in understanding your search?
How engaging is this view of a station?
How useful are the historic texts that describe this location?

The survey was comprised of questions like those in Table 9. There were two types of questions asked of users, quantitative answers, that asked for feedback on a scale from 1-5, and, qualitative answers, that allowed free responses using a text block. The survey

[1] https://github.com/areinhold17/deepmap

was open for several weeks and had over 60 unique users participated. Given that not all users accessed all questions, data from Google Analytics[2] was collected in order to monitor what pages users accessed and how long they used each page. Google Analytics is a web service that can be embedded into a web application, that records the details of users visiting the application. Google Analytics records things like the number of users actively using the application at any given time, the number of times a page was accessed by users, where in the world users are accessing the application from, and what devices users used to access the application. Not all of the data available from Google Analytics is relevant to the evaluation of the prototypes. The key datasets taken from Google Analytics for the evaluation were the number of users that accessed the prototypes during the survey period, and the number of times each page was viewed by users during that period. Google Analytics shows that there was a total number of 203 pageviews across the four prototypes, with all pages receiving at least 14 pageviews. This data shows that all parts of the prototypes had at least some level of visibility to the survey audience.

With the survey being linked to user interaction, there was limited feedback on some portions of the survey. The weakness of this method of evaluation can be seen in these limited areas of feedback. If these processes were repeated in future work, either the survey should be more rigid and not as linked to user interaction, or a greater number of participants, spending more time, would be needed to produce more usable results. Due to this limitation in the conducted survey, the case study evaluation sections (CSPs and CSOs), in some cases, are backed up by limited data that does not display well in a statistical representation, so those sections use the limited qualitative responses and present more as anecdotal.

The results of the survey were used in conjunction with other feedback to complete the evaluation rubrics for each case study, discussed below. While the survey results were helpful in completing the rubrics, the qualitative responses did not provide particular meaningful results. Most simply agreed with the relevant question and didn't provide significant opinions.

1.6.4 Evaluation Rubric

In order to compare the prototype applications to one another an evaluation rubric was designed that could help quantify how well each prototype addressed the definitions of Deep Mapping and the values and methods of Spatial and Digital Humanities. The rubric also includes criteria based on the research questions of this thesis, so that comparisons could be made between them. Table 2 shows a summary of each case studies rubric. In that table most of the criteria is met by all of the

[2] https://analytics.google.com/analytics/web/

prototypes. The criteria from the Spatial Humanities and Deep Mapping categories in the rubric that focus on time are the only ones not fulfilled by all prototypes (the red cells in Table 2). This discrepancy demonstrates that, while the ability to spatially navigate the data is important for creating a Deep Mapping application, a focus on the temporal nature of the data is not always a relevant element, depending on the intended design objectives of the application. The potential lack of temporal interaction does not eliminate these prototypes from being Deep Maps because many of the definitions presented by research of Deep Mapping do not have a focus or requirement of temporal interaction. Elements like spatial navigation, multimedia and layered design are criteria that appear in most definitions of Deep Mapping and those elements are present in all four of the prototypes.

The inclusion of a 'Maybe' answer for the fulfilment of criteria was because there were instances where the prototype did address some part of the criteria but not as completely as the others. This can be seen in the Crosthwaite maps response to 'Relationship between space and time.' There are references to temporal elements, specifically in the tabs for switching between text excerpts on the Station View (section 3.4), but they were a minor form of navigation and did not create a strong connection between the data being presented and its temporal context. In comparison the timeline on the Analytic Deep Map was a central feature for filtering the data and allowed the data returned by searches to be presented in a temporal context.

Overall, each prototype fulfilled the majority of the criteria in the evaluation rubric. The best performing case study was the Tarn Hows Deep Map which met all the criteria. This was somewhat surprising as the Tarn Hows application is the most abstract of the four prototypes. The worst performing prototype, by only one category, was the Reading in Place Deep Map. This was not as surprising given that the Reading in Place application had one of the simplest and most focused implementations, even though it included access to the full data set from the CLDW.

Table 2 Summary of Evaluation Rubrics from Case Studies

	Crosthwaite	**Analytic**	**Reading in Place**	**Tarn Hows**
Digital Humanities Criteria				
Iterative Process	Yes	Yes	Yes	Yes
Experimental	Yes	Yes	Yes	Yes
Collaborative	Yes	Yes	Yes	Yes
Multimodal	Yes	Maybe	Maybe	Yes
Open and accessible	Yes	Yes	Yes	Yes
Spatial Humanities Criteria				
Relationship between space and time	Maybe	Yes	No	Yes

Place built through multi-layered data	Yes	Yes	Yes	Yes
Deep Mapping Criteria				
Spatial Exploration as a primary form of interaction	Yes	Yes	Yes	Yes
Temporal context as a primary form of discovery	No	Yes	No	Yes
Data from verifiable source for legitimacy	Yes	Yes	Yes	Yes
Layers fill targeted area	Yes	Yes	Yes	Yes
Accessible to broad audience	Yes	Yes	Yes	Yes
Research Question Criteria				
Inform Research Process	Yes	Yes	Yes	Yes
Accessible and understandable				
Researchers and Scholars	Yes	Yes	Yes	Yes
Site Authorities	Yes	Yes	Yes	Yes
General Public	Yes	Yes	Maybe	Yes
Engaging				
Researchers and Scholars	Yes	Yes	Yes	Yes
Site Authorities	Yes	Yes	Yes	Yes
General Public	Yes	Yes	Yes	Yes
Repeatable	Maybe	Yes	Yes	Yes
Expandable	Yes	Yes	Yes	Yes

The Evaluation Rubric, seen in sections 3.6.3, 4.6.3,5.5.3, and 6.6.3, and summarized in Table 2 is a central element to the evaluation of future Deep Maps. The first three sections of the rubric, Digital Humanities, Spatial Humanities, and Deep Mapping are common criteria that can be used in the evaluation of Deep Maps in all forms. While not all criteria in these categories are essential to calling an application a Deep Map, they all contribute to understanding of how well an application fits within the various definitions of Deep Mapping and the wider research concepts that inform the Deep Mapping concept. An application that does not meet the majority of these criteria would be unlikely to be of sufficient design to be defined as a Deep Map.

The Research Question category of the rubric is heavily targeted at the interests of this thesis but are applicable criteria for most implementations of Deep Mapping. Access and engagement are important aspects of a Deep Map but not all Deep Mapping applications would be aimed at the same user groups defined by this thesis. Concepts like repeatability and expandability are potentially critical elements of any research project, not just the development of Deep Maps.

The Evaluation Rubric provides a tool by which applications can be evaluated to determine if they meet the requirements of being a Deep Map. The rubric also provides a template for comparing Deep Maps, even those that appear in different forms. Fulfilling criteria in the rubric can show common elements or themes between Deep Maps focusing on different methods of visualization or interaction, or on different regions or subjects of focus.

1.7 Thesis Structure

Chapter two describes the literature related to this thesis. This chapter will also cover background information on the technologies and methodologies used to develop the Deep Map prototypes created for this study, such as Natural Language Processing and Geoparsing.

The following four chapters discuss the design, development, and evaluation of the prototype Deep Maps created for this thesis. The prototype in chapter three is a Deep Map focusing on Thomas West's picturesque stations around Derwent Water Lake and their representation in Peter Crosthwaite's historic map.

Chapter four describes the second prototype developed, a Deep Map using the full corpus of Lake District texts collected for this project. This prototype used corpus linguistic techniques to search the corpus and map the results.

Chapter five describes another Deep Map using the full corpus of Lake District texts. This Deep Map focuses on close reading and engaging users with a single text. It allows for spatial exploration and contextualization while users read a text.

Chapter six discusses the last prototype developed through this research. This prototype is a 3D visualization of historic landscapes at the site of Tarn Hows, a significant heritage site and popular tourist destination in the Lake District.

Chapter seven discusses the last case study done for this thesis, not based on the creation of a prototype Deep Map. This chapter focuses on the processing of data not related to the Lake District, in order to determine if the methods used for the previous case studies are repeatable for other datasets.

Chapter eight summarizes the results of this thesis and discusses the potential for future work.

2 Literature Review

This section looks at the theories and methods that helped to inform the research undertaken for this thesis. This section includes background information on the projects that this PhD was a part of. It additionally contains background on the Lake District region, that is the main subject of this research, and background on the technologies used in the development of prototype applications created for the case studies.

2.1 Theory

2.1.1 Spatial Humanities

Scholars in the humanities have become increasingly aware of the impact geographic information can make in research. With the 'spatial turn' in the humanities, scholars have expanded research projects to include core technologies like geographic information systems (GIS) and web mapping in order to address spatial relationships in humanities data (Ayers 2010). The democratization of geographic information technologies offered new tools to humanities research for analysis of the spatial components of their data. Spatial Humanities research is usually done by interdisciplinary teams, making use of methods and theories from other disciplines that can answer questions that were impossible to address before (Murrieta-Flores 2019). Spatial Humanities makes heavy use of method and theories in geography for the mapping and analysis of data. In order to bring traditional humanities data into geographic systems, an initial stage of preprocessing is often required, drawing on technologies from disciplines like corpus linguistics and computing to convert humanities data into a machine readable and machine understandable format.

Space, Place, Time

This shift to spatial thought has focused on the concepts of Space, Place and Time. The standard uses of maps and related graphic methods have often focused on wayfinding and depicting geographic features and political boundaries, effectively showing Space, without attached meaning or significance. Space is somewhat of an abstract concept of a physical location that has not yet to be associated with personal experience. Place, in contrast with space, is a location that has attached meaning, usually through personal experience and memory (Porter 2018).

Meaning can be dependent on who has a relationship with a space. Locals from a small town might have significant meaning associated with the town given the memories, but individual and shared, that they have about it. Those locals would see the town as a place with meaning connected to many locations. Others, who have never been to that town, might only know of it as a location on a map, or a stop on the train, etc. and see it as a Space with little or no associated meaning (Tuan 1979).

Space is a setting for social and political expressions that create a unique Place. Historic actions and the development of culture embed stories in Spaces to create Place. "These stories are both individual and collective, and each of them link geography (space) and history (time)(Bodenhamer 2015)."

The concept of Time is important in the humanities study of Space vs Place. Time is one lens through which changes to Spaces and Places can be seen (Thrift 2006). A Space can become a Place over time as personal and societal experiences create meaning associated with the location. This type of change from Space to Place can be seen in places like the Lake District where locations that were once unexplored and untouched spaces become places that travel writers and tourists seek out and create a shared sense of meaning.

The same can happen in reverse, a Place can become a Space over time, as the associated meanings with that location are lost over time, due to changes in cultural practices or loss of human occupation (Thrift 2006). This is a common change seen in the archaeological record, where a location once had human occupation and thus some associated meaning, has subsequently been unoccupied and has lost its identity of place.

Mapping Projects in the Spatial Humanities

While research in the Spatial Humanities, by definition, have a spatial component, not all projects use map. Spatial analysis of data can occur without visual representation, like the methods described in section 2.5. Those projects that do include maps, can use maps of various depths to explore and present relevant data. The most common maps used in research are often traditional representations of maps, showing a cartographic area and a curated set of data related to the research being explored. These common maps don't tend to have the depth of data or level of interaction required to be a Deep Map. There are, however, some research projects that have created maps of greater depth that haven not specifically be classified as Deep Maps but do have the potential to fall within the broad definition of a Deep Map. The following examples are of research projects in the spatial humanities that could be classified as Deep Maps but are not explicitly described as such.

The Los Angeles Landscape History project create a set of interactive maps exploring the historic ecology of the Los Angeles River watershed and local environments. The maps created for this project incorporate ecological data from numerous historic sources, including indigenous and modern datasets. This project is the first example of a publicly available geo-historical database of spatial data to help enhance collaborative efforts for researchers interested in historic ecology of the region. The database contains map-layers, historic images, and textural records. The researchers collected data from diverse sources in order to cover 2.5 million years of ecological change in the Los Angeles area. To provide viable predictions of the ecological area and cover the intended timespan, the area was divided into a 1km square grid, and a set of predefined ecological classifications, such as grasslands and flower fields, foothill and valley forests, and salt marshes, were used. Each square of the grid could then be assigned a classification for a given time period based on the historic and modern scientific sources collected by the team. In addition to the publicly accessible database, the project team created several interactive maps to display the collected data. These maps are interactive, and each map displays a specific topic from the database. Since each map only displays one type of data, like ‘potential natural vegetation’, and the data displayed does not attempt to address Place as described above, these maps do not qualify as Deep Maps (Ethington 2020).

The Holocaust Geographies project has created several maps looking at World War II Jewish Ghettos, Concentration and Labor camps, and other topics related to the experience of Jewish people during and after the war (https://holocaustgeographies.org/). This project makes significant use of GIS to map historic data, and also uses the research environment as a platform to critique GIS as a tool in the humanities. The large GIS applications, like ArcGIS from ESRI, are available to all disciplines, but they are typically built for more tradition geospatial disciplines like geology, hydrology, and urban planning. While GIS is still an powerful tool in the humanities and can have significant bearing on research, the Holocaust Geographies project looked at GIS from a more critical viewpoint in order to understand how traditional GIS impacts research and what it lacks when addressing research questions in the humanities (Westerveld 2020).

The data for project maps comes from a variety of source, including historic maps and survivor testimonies. The survivor testimonies were processed as part of this PhD in order to make them mappable, as described in section 6. The ‘Budapest Ghetto dataset’ map created by the project includes multiple layers of data, such as Jewish-designated residences, Ghetto building and areas, and Locations where actions took place, like streets being cleared of Jews. Locations in the map can be clicked on to bring up a window with additional information. This map is a good example of incorporating multiple historic sources of spatial information into a unified map. The map has multiple layers of data and has some user interaction to explore the given area. This map could be considered a Deep Map, but the interaction with the data and layers by users is limited, which would make this map partially meet the criteria for a Deep Map (Knowles 2020).

The other maps in created by the Holocaust Geographies project also have the potential to be Deep Maps. They each present multi layered data to users with some level of interactive behavior. Two of the maps, the 'Arrests in Italy dataset' and the 'Holocaust Survivors in Budapest dataset', use database queries to display data in the map, which requires users to input a name in order to explore the data. The requirement to input a name doesn't make the maps an open-ended exploration, since some prior knowledge or random guessing is need. This makes these maps partially lack some of the criteria for a Deep Map, but they are excellent exemplars of interactive maps in the spatial humanities.

2.1.2 Digital Humanities

Spatial Humanities is often linked with Digital Humanities. Spatial Humanities makes significant use of digital technologies like GIS in order to assist research, and in this way overlaps with digital humanities, but does not explicitly require a digital approach. Digital Humanities (DH) is a field of research at the intersection of digital technologies, usually software driven solutions, and the various fields of humanities, like history and literature. The field of DH is diverse given its intersection between the humanities and more hard science fields like computer science, but a set of values and methods were proposed to help to unify research with common threads (Honn 2015). Five values were proposed as key elements to DH research.

"Critical and theoretical"

DH like most scholarly practices is based in theory and critical reasoning. DH draws theoretical methods and practices from various disciplines in the hard sciences and humanities to enable new approaches to research. The humanities side of DH provides grounds for self-criticism of the theories, methods, tools, and technologies used (Honn 2015).

"Iterative and experimental"

Like many software development projects and computer science research, DH makes use of iterative versioning. DH projects continuously build upon the knowledge gained throughout the project. This iteration can be seen in the development of the Spatial Humanities and Geospatial Humanities projects that this PhD research is a part of. The iteration in DH provides an environment for 'experimentation, risk-taking, redefinition, and sometimes failure (Lunenfeld et al 2012).'

"Collaborative and distributed"

DH projects are almost entirely multidisciplinary, with contributors from differing fields, with varying experience and expertise, that bring with them theories, methodologies, and processes that might otherwise never be considered for application to a project's research goals (Honn 2015).

"Multimodal and performative"

Being part of the digital environment, DH embraces flexibility in the publication and presentation of research. Varying from traditional text in print to audio and video, digital multimedia products allow for augments to be expressed in optimal modes and visualizations (Honn 2015).

"Open and accessible"

DH often embraces the openness of the digital world and makes use of existing opensource resources, and where possible contributes to them. Tools like GitHub and Wikipedia provide public forums for sharing knowledge and enables continued collaboration within the community. This collaboration allows for DH projects to continue impacting the field beyond the scope of the project and enables some of the iterative process that DH values (Honn 2015).

Methods:

The methods in DH vary significantly depending upon the scope of projects and the make up of the interdisciplinary teams. The methods used in DH are closely related to but not explicitly linked with the technologies used in DH research. The book *Digital Humanities* (2012) discusses a collection of common methods that are used within DH, aggregated by Josh Honn (2015) below:

> ***Enhanced Critical Curation***
> *Object-based arguments through the curation of digital media, including collection repositories and scholarly narratives supported by digitized or born-digital primary source materials.*
>
> ***Augmented Editions and Fluid Textuality***
> *Digital critical editions, marked up and encoded texts, often created*

through crowd-sourced methods and open to perpetual revision, annotation, and remix.

Scale: The Law of Large Numbers
As data sets grow larger and larger, humanists hope to create new findings through computational- and algorithmic-enabled interpretations of our digitized and born-digital culture materials.

Distant/Close, Macro/Micro, Surface/Depth
In contrast to, and often in conjunction with, close reading, distant reading looks to understand and analyze large corpora across time through "trends, patterns, and relationships."

Cultural Analytics, Aggregation, and Data-Mining
Through computational means, cultural analytics mines, studies, and displays cultural materials in new aggregated or remixed forms, often including interactive and narrativized visualizations

Visualization and Data Design
Arguments made from the visualization of data, including virtual/spatial representations, geo-referencing and mapping, simulated environments, and other designs constructed from and informed by data.

Locative Investigation and Thick Mapping
The creation of "data landscapes" through connecting real, virtual, and interpretive sites, often manifesting as digital cultural mapping or geographic information systems (GIS).

The Animated Archive
In which the static archive of the past is made alive and virtually experiential, including the active archiving of physical spaces through virtual means, and multi-modal/faceted approaches to collection access and interactivity.

Distributed Knowledge Production and Performative Access
Digital projects take collaborative teams that cross both disciplines and borders and that often challenge the idea of "the author" through team contributions, crowdsourcing, and the user-based performance of the "text."

Humanities Gaming
Taking on "historical simulation," humanities gaming uses virtual

learning environments to create interactive narratives that engage users and enable the exploration of humanist themes.

Code, Software, and Platform Studies
Humanists have studied texts, the book, and many other forms of writing, so what to make of the code programmers write, the software computer users use, and the platforms that shape our social and cultural interactions?

Database Documentaries
Multi-modal narratives formed from a database, branching out into multiple paths users explore, possibly incorporating live-feed data, all calling into question authorial control/intent and the role of the reader.

Repurposable Content and Remix Culture
Digital content can be read, written, and rewritten, and as such all digital objects are subject to sample, migration, translation, remix, and other forms of critical reuse.

Pervasive Infrastructure
Our digital realities encompass many types of machines and screens and increasingly our objects are stored in the cloud, distributed over servers in multiple locations, so what does that mean for humans and data?

Ubiquitous Scholarship
Print publication no longer is the only way forward, and as new modes of publishing proliferate, and new players in publishing participate, publishing becomes increasingly ubiquitous and open.

Many of these methods are applied to the research undertaken in this PhD. Most of the case studies have some focus on Visualization and Data design methods, and Locative Investigation and Thick Mapping. The NLP Deep Map, chapter 4, and Close Reading Deep Map, chapter 5, approach the Distant/Close reading dynamic. Chapter 6, Tarn House Deep Map, focuses on virtual environments which related to the Humanities Gaming methodology.

2.2 Geospatial Innovations Project

The 'Geospatial Innovations in the Digital Humanities' project[3], which this research is a part of, is aimed at expanding the research done by two previous projects at Lancaster University. The 'Mapping the Lakes' project[4], undertaken at Lancaster University between 2005 and 2007, was the first of these projects. The second project was the 'Spatial Humanities: Text, Place, GIS'[5], undertaken between 2012 and 2016.

2.2.1 Mapping the Lakes

'Mapping the Lakes' focused on two journeys of the poets Thomas West (1769) and Taylor Coleridge (1802), in the Lake District region. This project used the texts of West's and Coleridge's journeys and identified each mention of a location in the text and assigned metadata to each, including whether it was visited by the author or just mentioned, the order of visitation, and the geographic coordinates. This information was then brought into GIS to map the journeys against each other. This spatial comparison highlights the differences and similarities between the two authors journeys (Gregory 2009).

The goal of this project was to test whether Geographic Information Systems (GIS) could be used to expand spatial thinking of literature focusing on place and space. This project used interdisciplinary methodologies to apply geographic and spatial technologies to consider historic and literary problems. The first stage of the project was to manually identify the place names present in the two texts and create an Access database of the locations, noting a distinction between the locations visited by the authors and those mentioned but not visited. The locations were then manually associated with the EDINA Digimap Gazetteer[6], and the list of locations and coordinates were used to create a GIS map.

The second stage of this project was to use the GIS map to explore research questions related to space. The map was used to address questions like the following:

- How did the two writers define the boundaries of the English Lake District, whether real or imaginative?
- Did the writers move outside of the central Lake region? What areas did they visit or mention that are not commonly explored?
- Did the writers use similar language when discussing the similar places?

[3] http://wp.lancs.ac.uk/lakesdeepmap/
[4] https://www.lancaster.ac.uk/mappingthelakes/
[5] https://www.lancaster.ac.uk/fass/projects/spatialhum.wordpress/
[6] https://digimap.edina.ac.uk/

The GIS map was used with various visualization methods to display the locations and visitation data so that the spatial elements of the two texts could be used to address these questions (Cooper 2011).

2.2.2 Spatial Humanities

The 'Spatial Humanities: Text, Place, GIS' project focused on expanding the research done by the 'Mapping the Lakes' project. The project created a large corpus of historical texts, the Corpus of Lake District Writing (CLDW), including books, periodicals and official reports (Donaldson 2017). This corpus contains 80 manually digitized texts about the Lake District region from 1622 to 1900. The corpus is made up of several different genres, including travel journals, essays, novels, and poetry. The corpus includes many key writers from the time period, mostly male, although six of the texts are written by women. The authors are primarily from the UK but also represent other nationalities, including the United States and Germany. The corpus contains over 1.5-million-word tokens, across the 80 texts. The term 'corpus' and the field of corpus linguistics is defined below in section 2.5.1

The Spatial Humanities project processed the CLDW's 80 texts through the Edinburgh Geoparser (Grover 2010), explained in detail in section 2.5.3. Using the Edinburgh Geoparser, locations mentioned in the text were identified and annotated with geographic coordinates. This annotation took the form of 'enamex' eXtensible Markup Language (XML) tags, which include attributes for latitude and longitude, place name, confidence score, and source gazetteer. In addition to this annotation, 28 of the 80 texts in the corpus were manually annotated, to create a Gold Standard, in order to evaluate the performance of the geoparsers results. The Gold Standard texts were selected to represent the various genres and historic periods present in the CLDW (Table 1). This subset consists of approximately 242,000 word tokens, making it about one sixth of the total corpus. Each text in the Gold Standard subset was reviewed by hand, and each place name instance was identified and tagged with a 'cdplace' tag. This annotation can be seen in an excerpt from "An Extract of the Rev. Mr. John Wesley's Journal":

> *<p>IN the Evening I preached at <cdplace><i>Bolton,</i></cdplace> and on <i>Friday,</i> 11. about Nine at <cdplace><i>Lower Darwent,</i></cdplace> a ßmall Village near <cdplace><i>Blackburn.</i></cdplace> At <cdplace><i>Lancaßter</i></cdplace> we were informed, it was too late to croßs the Sands. However we reßolved to make the Trial. We paßßed the Seven-mile Sand without Difficulty, and reached <cdplace><i>Fluckborough</i></cdplace> about Sunßet.</p>*

TABLE 3 GOLD STANDARD SAMPLE

AUTHOR	TITLE	DATE	WORD TOKENS
LT. HAMMOND	A Relation of a Short Survey of 26 Counties	1634	2,492
DANIEL DEFOE	A Tour thro' the Whole Island of Great Britain	1725	3,019
GEORGE SMITH	'Survey of the Northwest Coast of England'	1746	2,172
GEORGE SMITH	'A Journey up to Cross-Fell Mountain'	1747	2,774
GEORGE SMITH	'A Journey to Caudbec Fells'	1747	6,955
JOHN BROWN	Description of the Lake and Vale at Keswick	1751	1,107
JOHN WESLEY	An Extract of the Rev. Mr. John Wesley's Journal	1759	1,783
THOMAS GRAY	Journal of A Visit to the Lake District in 1769	1769	6,388
THOMAS PENNANT	A Tour in Scotland. MDCCLXIX	1769	1,069
ARTHUR YOUNG	Six Months' Tour Through the North of England	1770	13,388
THOMAS PENNANT	A Tour in Scotland, and Voyage to the Hebrides	1772	13,994
THOMAS WEST	A Guide to the Lakes	1778	36,793
WILLIAM COCKIN	Ode to the Genius of the Lakes. A Poem	1780	4,793
R. J. SULLIVAN	Observations Made During a Tour	1780	2,799
STEBBING SHAW	A Tour in 1787	1787	7,073
J. H. MANNERS	Journal of a Tour	1796	1,311
S. T. COLERIDGE	Letters and Notebooks	1802	9,237
PRISCILLA WAKEFIELD	A Family Tour through the British Empire	1804	5,459
ANONYMOUS	Gleanings of a Wanderer	1805	1,802
JOHN KEATS	Letters	1818	3,794
WILLIAM WORDSWORTH	The River Duddon: A Series of Sonnets	1820	10,938
JONATHAN OTLEY	A Concise Description of the English Lakes	1823	29,704
JOHN RUSKIN	Iteriad; or, Three Weeks Among the Lakes	1830	21,346
WILLIAM WORDSWORTH	Guide through the District of the Lakes	1835	5,464

ANONYMOUS	Keswick and its Neighbourhood: A Handbook	1852	29,494
JAMES F. CLARKE	Eleven Weeks in Europe	1852	1,606
ANONYMOUS	The English Lakes	1859	10,126
HERBERT RIX	'Down the Duddon with Wordsworth'	1893	5,412
		Total	242,292

To test the potential of the corpus as a gold standard, two standard NLP libraries were used, the Stanford CoreNLP[7] and OpenNLP[8] libraries. These libraries were used to process the corpus with NER to detect place names. A custom geoparsing engine was built as part of this thesis that could take the results from the NLP processing. The geoparser allows one or more gazetteers to be added to the system, for this test of the corpus we used the GeoNames[9] gazetteer of the UK as a baseline. The geoparser looks at each place name and identifies every potential match in the gazetteer using three types of matches, exact match, inner match, and outer match. In an exact match, the two terms being compared are identical, though not case sensitive. The inner match matches the place names term with part of the gazetteer entry (Place name "Britain" would match to "Great Britain"). An outer match is the reserve of the inner match (Place name "Great Britain" would match to "Britain"). Each match type is given a confidence rating based on the Levenshtein distance calculation [LEVEN], which calculates the number of changes between two terms (exact match = 0, "Rydal" -> "Rydal Mount" = 6). In addition to the confidence rating, the geoparser puts a priority on locations closer to the Lake District, since the corpus is focused on that region (Rayson 2017).

After processing the corpus through each NLP library and the geoparser, a comparison tool was created to compare the golden standard geoparsed text to the newly processed text. The comparison tool looks at how many locations were identified in the golden standard and from each NLP library, then it looks at the number of locations Identified as the same locations, and the amount identified as a different location. The results of this comparison across the whole corpus shows that the OpenNLP library identified the fewest place names in the texts, and the golden standard identified the most place names. The Stanford library and Geonames Gazetteer matched 49% of the identified locations with the equivalent location in the golden standard (Rayson 2017).

[7] https://stanfordnlp.github.io/CoreNLP/

[8] http://opennlp.apache.org/

[9] https://www.geonames.org/

This baseline comparison of the golden standard corpus to two NLP libraries and Geonames gazetteer shows that the processing and manual correction of the corpus had produced a solid golden standard. While the processing of the corpus through the Edinburgh geoparser had left a few false positives, the manual correction of the data has caught many of these false positives and corrected inaccurate gazetteers matches. By adding historic gazetteers to our geoparser, the results of place name geolocation should improve in future processing and increase the match score above 49%. Increasing the accuracy of geoparsing in the corpus will allow for more fine-grained analysis of spatial relationships and allow for a more complete representation of the spatial distributions present in the texts. With a historic gazetteer and greater accuracy in processing the more obscure locations described in the texts will become more readily available for corpus linguistic analysis and integration into spatial applications like Deep Maps (Rayson 2017).

2.2.3 Geospatial Innovations

The Geospatial Innovations project[10] is funded by the Leverhulme Trust. The project's goal is the development of new understandings of the literary and cultural geographies of one of Britain's most significant cultural landscapes, the English Lake District, by applying groundbreaking, exploratory geographical methods to the interdisciplinary research field of the spatial humanities. Bringing together researchers with complementary expertise in computer science, geographic information science (GISc), literary studies, and regional history, the project is aimed at changing the way scholars engage with the geographies that inform regional identity and sense of place. The stated objective of the Geospatial Innovations project are as follows from the final report:

1. Design and develop a prototype deep map that will lead to a step-change in the digital representation, exploration and analysis of place and space in regional history and literary studies.
2. Generate new insights about the cultural geography and literature of the Lake District drawing on a large and diverse range of source materials.
3. Complete two book-length scholarly works that use contrasting, but complementary, approaches to exemplify how the materials from this resource can make specific knowledge contributions.
4. Increase the use of geospatial technologies within and outside the academy by establishing a research network in geospatial cultural and literary studies.

[10] http://wp.lancs.ac.uk/lakesdeepmap/

This thesis concentrates on addressing the first objective of the Geospatial Innovations project, in the development of Deep Map prototypes. The Geospatial Innovations project concluded in 2018, with the exception of this PhD research which started mid way through the project. The project achieved its aims of contributing to the understanding of literature geographies. Several books and numerous papers were published as part of this project to demonstrate what new, technically based digital humanities approaches can offer to literary scholars who are not necessarily interested in digital methods.

2.3 Lake District National Park

The Lake District National Park is the primary subject of this thesis's research, with an emphasis on historic literary data sources, given the significant literary history of Lake District region. The Lake District National Park is the UKs largest national park[11]. Located in the northwest of England, the Lake District is a natural landscape whose iconic valleys were carved out by glaciers and later shaped by agricultural and pastoral land use. The Park spans 912 square miles and is known for its landscape of high mountain peaks and many glacial lakes. There are over 200 mountains, or fell, tops in the region, with the highest reaching 978 meters above sea level. The region contains numerous large lakes, like the largest, Windermere at just under 15 square miles, and many Tarns. Tarns are small mountain lakes or pools.

The Lake District was made a UNESCO World Heritage site in 2017[12]. The district is recognized on the World heritage list for its natural and cultural significance. The region is filled with pastoral fields enclosed by stacked stone walls, which delineated the land after the General Enclosure Act of 1801. Human occupation of the region dates back over 12000 years. Significant periods of the region include the Roman Occupation, Medieval era, Industrial Revolution and the Victorian era. During the 18th century the Lake District became a hub of literary thought and the dynamic landscapes helped inspire writers in the Picturesque and Romantic movements.

In modern times the Lake District has been classified as a national park and a number of significant historical sites are owned and maintained by the National Trust[13]. Many of these historic sites owned by the National Trust relate to well known authors

[11] https://www.lakedistrict.gov.uk/home

[12] https://whc.unesco.org/en/list/422/

[13] https://www.nationaltrust.org.uk/days-out/regionnorthwest/lake-district

included in the CLDW, including the Wordsworth House and considerable amounts of land owned by Beatrix Potter.

2.4 Deep Mapping

Deep Mapping is a somewhat ambiguous term that appears in various disciplines. This thesis and the proceeding review focus on Deep Mapping in the context of the Spatial and Digital Humanities. The concept of Deep Mapping emerged in literature in the 1990s but evolved in the last ten years to become an emerging concept in the Spatial Humanities.

2.4.1 History

The Spatial Turn in the humanities in recent years has increased scholar's awareness of spatial data and geographic information systems as necessary tools for research. Maps are a familiar cartographic product that effectively represent specific features within a spatial context. Geographic Information Systems (GIS) have been a key technology in the production of maps. The computational power of GIS has allowed for greater scope when building data driven thematic maps (Oxx 2013). The use of GIS in the humanities enables new methodologies to be incorporated into research. Even with the new research potential, the humanities have been slow to adopt GIS. This pace may be explained by the requirement of a significant amount of expertise to plan, create and manage GIS projects. Additionally, humanities research and the data sources associated with projects is often difficult to translate to a spatial framework. Extracting spatial data from historic text-based sources, for example, can prove difficult, as such operations require another layer of processing and expertise to produce the quantitative data that GIS requires (Alves 2015). GIS are versatile in their ability to incorporate data, but it has been recognized that they are lacking in the ''ability to handle temporal dimension and facilitate spatio-temporal analysis (Oxx 2013)". An attempt, in the discipline of history, has been made to incorporate time as a variable into historical geographic information systems (HGIS). A temporal variable allows for data to be joined with other common GIS variables using spatial attributes. As temporal variables are not standard within GIS, joining data based on temporal attributes can create relationships that "tend to multiply and generate datasets that contain more exceptions than rules, more arbitrariness than standards (Silveira 2014 p.60)".

With the spatial turn in the humanities and the increasing need for spatial applications that move beyond the limitations of GIS, a new focus has been made on the "re-conceptualization and exploration of maps as 'Deep Maps'" (Oxx 2013). The concept of Deep Maps was introduced by William Least Heat-Moon in *PrairyErth* (Heat 1991). Heat-Moon describes a Deep Map as a spatial narrative. *PrairyErth* is a narrative of Chase County, Kansas, and focuses on the spatial elements of the cultural traditions of the Plains. He describes "rock art, elk skin celestial charts, birch bark scrolls, and other forms of indigenous map making," in an effort to record the diverse spatial traditions and knowledge of the region and its native people (Maher 2005).

Shelley Fishkin uses 'Deep Maps' as an acronym for 'Digital Palimpsest[14] Mapping Projects (DPMPs)'. Fishkin points out that "there might be as many varieties or Deep Maps as there are scholars, and they might take a dizzying range of forms (2011)." Considering this, Deep Maps should endeavor to include a set of common features:

- Interactive Map with embedded links to texts and images from archival sources. A Deep Map should contain data that has a reasonable and verifiable source in order to provide legitimacy.
- Focus on topics that span the spatial context and include layers of data that fill the targeted area. All Maps inevitably have a limited scope, of the space in which they cover and the content that was chosen to be displayed, but a Deep Map should endeavor to provide data points that cover the whole extent of the area of interest.
- Accessible to a broad audience. A Deep Map should be open access, allowing anyone interested in the subject matter to easily access the application and content. Open access would additionally provide for access to the application on a global scale, and not be limited to regional use.

While Deep Mapping has its roots in literature and the humanities, a similar concept has been discussed in the digital humanities discipline. Thick mapping, a term described by Todd Presner, is a process "of collecting, aggregating, and visualizing ever more layers of geographic or place specific data." Presner links thick mapping with Deep Mapping because they both focus on temporal and historical dynamics through multi layered data. Thick mapping differs from Deep Mapping in that while Deep Mapping is essentially focusing on a vertical view of a place through time, thick mapping focuses on the horizontal view as well, taking a broad view of data available at a single time side by side with data over time (Presner 2014).

David Bodenhamer, one of the foremost authorities on the conceptualization of Deep Mapping in the spatial humanities, states deep mapping *"conflates oral testimony, anthology, memoir, biography, images, natural history and everything you might ever*

[14] A Palimpsest is "something such as a work of art that has many levels of meaning, types of style, etc. that build on each other," often referring to writing (https://dictionary.cambridge.org/us/).

want to say about a place, resulting in an eclectic work akin to eighteenth and nineteenth-century gazetteers and travel accounts. Its best form results in a subtle and multi-layered view of a small piece of the earth (Bodenhamer 2013, p.27).” This definition, like others, focuses on the layered design of Deep Maps and the inclusion of different multimedia data sources into a spatial system. Bodenhamer later defines a Deep Map as *“simultaneously a platform, a process, and a product. It is an environment embedded with tools to bring data into an explicit and direct relationship with space and time; it is a way to engage evidence within its spatiotemporal context and to trace paths of discovery that lead to a spatial narrative and ultimately a spatial argument; and it is the way we make visual the results of our spatially contingent inquiry and argument (Bodenhamer 2015 p.4).”* These definitions concentrate on Deep Maps as tools in the research process and dissemination of data. Like other definitions, Bodenhamer’s definition makes the connection between Deep Mapping and the Spatial Humanities ideas of Space, Place and Time.

2.4.2 Recent Work

The NEH Institute on Spatial Narrative and Deep Maps[15] held a workshop aimed at the development of models for Deep Mapping.

One group from the workshop built a prototype, using Google maps to spatially place data and a content bar to allow users to view details about visible sites, which is a good example of the basic implementation of a Deep Map, because it can provide information to a user that can transform a space into a place. The creators of this implementation of a Deep Map put the most ‘emphasis on one particular user interaction: exploration’ (Ridge 2013). This group’s prototype, while simple in its implementation, allowed for the team to explore the concept of Deep Mapping and conceptualize a model for the relationship between Spatial Narratives and Deep Maps, Figure 1.

[15] NEH Institute on Spatial Narrative and Deep Map, June 18 to 29, 2012 - https://www.neh.gov/divisions/odh/institutes/spatial-narratives-and-deep-maps-explorations-in-advanced-geo-spatial

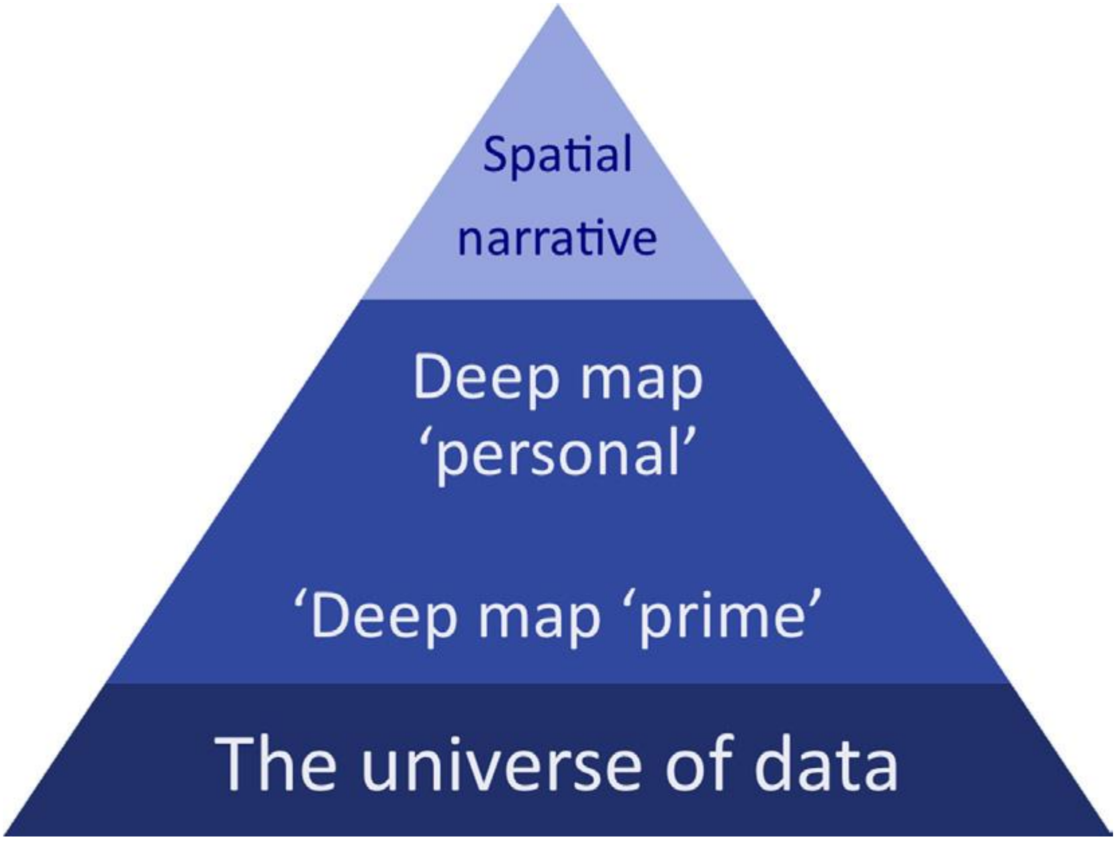

FIGURE 1. CONCEPTUAL MODEL OF THE RELATIONSHIP BETWEEN DEEP MAPS AND SPATIAL NARRATIVES (RIDGE 2013)

Spatial narratives are a curated representation of data, similar to a Deep Map, but they don't allow for the open-ended exploration of data, rather they are directed and guide users through the content. The inclusion of two different levels of Deep Maps in this model is intended to demonstrate that Deep Maps can be focused or broad. A focused map, or 'personal' Deep Map, is a partially curated collection of data in an explorable, spatial format. They can also be a more 'greedy' application, or 'prime' Deep Map, attempting to include all available data. With the ubiquity of the internet and the increases in linked data, the creation of a 'prime' Deep Map is becoming more possible, but attempting to include all possible data, the 'universe of data,' would be very difficult, given the current limitations of technology. 'Personal' Deep Maps, on the other hand, are well within the capabilities of modernly available technologies.

Another group from the workshop focused on using data from Greek Orthodox churches in Indianapolis, creating a Deep Map that focused on the 'spatial navigation' of data, rather than the spatial distribution of data (Oxx 2013). Instead of placing all the data they had available into a map, this team created a Prezi project and placed their evidence into the canvas of the presentation. Each 'slide' in the presentation contains some curated data, ranging from text, to images, to geographic maps. The slides themselves are spread across the canvas of the application in a spatial environment. This structure allows the data to be spatially navigable by viewers of the presentation, but the data does not need to be associated with any geographic coordinate systems. The presentation was also given a storyline so users could be led through the data in a more curated manner, but users still have the freedom to explore and deviate from the storyline if they want. The Prezi project shows a more abstract view of what a Deep Map could be. This abstract interpretation of a Deep Map still meets the definitions described above but deviates from the other implementations that have often focused on interactive maps.

Outside of the NEH Workshop few other practical applications of Deep Mapping have been created. One of the most substantial applications created so far is for the Deep Maps: West Cork Coast Cultures project[16]. This project looks to explore the rich maritime environment of the Cork coast, focusing on Roaring Water Bay, a stretch from Timoleague to Bantry Bay. The project brings together the environmental and biological records available in the area and combines those with the user submitted local cultural histories ranging from the early 1700s to the 20th century.

The Deep Map created for the West Cork Coast Cultures project is an online ArcGIS Map[17] of the region. The map contains an overlay of a historic map in addition to the standard satellite imagery. The map also contains map markers divided into eight categories. The categories include Visual Culture, Poetry, Early 19th Century Accounts, 18th Century Accounts, Environment, Species, Soundbites, and Folklore. Each mark can be clicked on which opens an info window. The info window contains information relating to the location of the marker and the category it represents, ranging from textural data to multimedia like images and sound clips. The map also has a layer of polygon data depicting fishery information along the coast.

The Cork Deep Map is a relatively simple design. The content has a fair amount of breadth, ranging from biological and environmental records to folktales and person accounts, and from video/audio clips to images and text. While the content has breadth it doesn't necessarily have depth. Each marker on the map pops up a small info window with various content, but there are no connections between markers that may have common links or related content. Each location usually only has one entry. The historic map base layer is interesting for the rough context of historic locations but lacks resolution for detailed study of specific places.

This Deep Mapping example shows a different interpretation of Deep Mapping concepts. This map 'aims to renew our sense of the depth of meaning associated with places over time,' and does so by including content that has deeper meaning to the

[16] http://www.deepmapscork.ie/deep-map-west-cork/deep-maps/

[17] http://www.deepmapscork.ie/deep-map-west-cork/deep-map-west-cork/

various topics of the map like folklore and biology. The map doesn't focus as much on the meaning of a specific singular place over time, as much as it focuses on the region as a place.

Non digital Deep Maps

Deep Maps do not necessarily need to be digital. While the digital platform enabled greater degrees of interaction and quantities of data, there are examples of non-digital maps that can qualify as Deep Maps. An example of these non-digital Deep Maps can be seen in the Crosthwaite case study, section 3. The Crosthwaite map of Derwent Water is a traditional printed map, but the goal of the illustrator was to depict the picturesque location around the lake. The map of the lake has various descriptions, notes and drawings of views from locations around the lake that enhance a viewer's understanding of the area beyond the information present in the topographical map. The additional data gives viewers of the map the potential to connect with the places in the map without having to be there in the real world. This map is an excellent historic example of how multiple types of data can be brought together in a single representation to create a Deep Map.

Limitations

Deep Mapping is a tool that can enhance research and dissemination, but also has limitations based on available technologies. Deep Mapping in digital applications is limited by the capabilities of the modern software like web browsers, which have restrictions on memory usage and graphics. Digital mapping is also limited by data, requiring data to be digitized in a machine-readable format. This is an especially significant limitation in the context of historic research as most data is in analog formats that are labor intensive to bring into the digital space. As technology improves these limitations will be reduced. Technologies like Optical Character Recognition have already improved the ability of computers to ingest historic texts.

2.5 Technologies

The Primary data used in the creation of Deep Maps for this research is the Corpus of Lake District Writing. This collection of texts was processed, analyzed, and manipulated using the following technologies and methodologies in order to prepare the texts so they could be visualized in mapping applications.

2.5.1 Corpus Linguistics

Corpus Linguistics is the study of language using large sets of data. The term 'corpus linguistics' covers two types of intellectual thought (Tognini-Bonelli 2002). The first, 'corpus-based' studies, is a set of procedures and methods for the study of language within a collection of data in order to explore theories or hypotheses established in traditional literary studies. The second, 'corpus-driven' linguistics, looks at the collection of data independent from literary theories or hypotheses in order to act as a source of new ways to derive hypotheses from the data.

The large set of data used in both these is called a corpus, and while the data contained in a corpus is often machine-readable text. A corpus can be made up of other types, like paper writings (Fries 1952), or video records of gestures associated with language (Knight et al 2009). A corpus is usually, though not always, of such a size that it defies traditional analysis by manual human processes within a reasonable amount of time. In many studies corpora often contain over a million words within the collection of texts, though some studies have used much smaller more targeted collections of data to address specific research questions. There are two types of corpus that can be created for research. The first, a monitor corpus (Sinclair 1991), is a corpus that continually adds more texts over time, such as a corpus of modern newspapers or a corpus of twitter messages. The second, a balanced or sample corpus (Biber 1993 and Leech 2007), is a corpus that contains a sampling of texts from a specific point in time, place or subject. The CLDW is an example of a sample corpus because of its focus on literary works from the Lake District region between 1600s and 1900s.

The procedures and methods that are used in corpus linguistics are still developing as new ideas and technologies emerge for the analysis of corpora. Two of the most well established analytical methods are concordancing and word frequency lists. Concordancers are tools that allow for words to be viewed in context. A concordance is a list of the occurrences of a word and the words plus or minus some value on either side each occurrence of that word in the text. Table 2 shows an example of a concordance search on the word 'Derwent' in the CLDW.

TABLE 4 CONCORDANCE EXAMPLE

Concordance 10 to the left	Search Term	Concordance 10 to the right
and walked eight miles to Breakfast at Keswick on	**derwent**	water-- We could not mount Helvellyn for the mist
see the Falls of Low Dore -- The Approach to	**derwent**	water is rich and magnificent beyond any means of conception--
to take(I doubt not) an equally flying view of	**Derwent**water	Robin Partridge, when he told us of it, asked us

leaving it - Awoke by Dancing - A blind Fidler - 165 Chap.	**Derwent**er	Lake.- Went upon the Lake - A bottom Wind - Effect on
first saw it, it was" wondrous, nay passing fine," it	un**derwent**	several degrees of shade, what with the sun, and time,
Wad Mine-- Herbert's Island-- A Reflection on it-- The Lady	**Derwent**erwater's	Escape-- Manager's Speech on having a thin House-- A short

A word frequency list shows all the words that appear in a corpus or sub selection or texts and shows the quantity of occurrences, see Table 3. Concordances and word frequency lists allow for users to evaluate the qualitative and quantitative significance of words within the corpus.

TABLE 5 WORD FREQUENCY EXAMPLE

Word	*Frequency*
hithermost	1
shore	4
distance	10
heard	7
murmur	1
audible	3
day-time	1
wished	3

2.5.2 Natural Language Processing

While corpus linguistics allows for the analysis of a corpus to answer research questions and provide context for searches of words within the texts, the procedures and methods of corpus linguistics are typically applied to the surface word forms only and without further annotation cannot analyze the texts with an understanding of the grammar or syntax of the content. Natural Language Processing (NLP) is a Machine Learning (ML) approach to the processing of a corpus in a way that understands the grammar and syntax in order to derive the structure of sentences. By combining NLP and CL, we can annotate a corpus and then apply similar methods (concordancing and frequency lists) to analyze the annotation (Rayson, 2008).

Machine Learning is a term for computer algorithms that use statistics to find patterns in large amounts of data. Machine learning algorithms are designed to automatically learn from past experience (data previously processed through the algorithm) in order to better find patterns in the data and perform tasks associated with those identified patterns. Machine learning is similar to computational statistics which focuses on using data statistics to make predictions (Dutt 2018). Machine learning uses models to structure how the computer will process data to make predictions and identify patterns. These models vary depending upon the type of processing required given the desired tasks to perform with the data. Models are trained with a set known data relating to a task and the algorithms learn from that data in order to process additional data.

Natural Language Processing uses machine learning algorithms to understand the structure of sentences or syntax (Martinez 2010). The syntax helps in determining the meaning of a sentence, the semantics. NLP systems are usually machine learning algorithms that have been trained using a marked-up corpus of data. A marked-up corpus would have each word in the texts tagged with a semantic or other linguistically meaningful label so that the computer can understand what each word is in context. NLP uses a defined set of rules for the grammar of a language. Context-free grammars (CFGs) are the most common type of grammar used in NLP applications. CFGs define elements within text, like terminal symbols (punctuation and words), that provide the basic structure of language (Martinez 2010).

NLP uses ML algorithms to solve issues like word sense disambiguation (Diamantini et al), when words have different meanings depending on context. ML algorithms trained on tagged data are used to assign which meaning a should be attached to a word in each particular use of it (Martinez 2010). Identifying the meaning of a word in context helps with further processes like identifying the part-of-speech a word is, since some words can be more than one depending on context. Examples of words with multiple meanings or multiple parts of speech can be seen in these sentences:

TABLE 6 WORD SENSE DISAMBIGUATION

Change in meaning	Change in meaning and part of speech
That bird is a *crane*.	The *object* is on the table.
They had to use a *crane* to lift the object.	I *object* to those terms.

Part of speech (POS) tagging is an aspect of NLP required to accurately parse text (Charniak 1996). POS tagging assigns words to their syntactic categories, such as noun or verb, based on the context of that word in the sentence. POS tagging is usually performed using Markov chains or hidden Markov chains (Manning 2000), which are ML models that use a mathematical system to transition from one state to another. Markov chains have a fixed set of states for which a process can end, in this case those states are the various syntactic categories. The model used the tagged corpus to learn the probabilities of POS tags appearing in sequences or patterns, such as the basic grammatical word order of: subject, object, verb. Like many aspects of NLP, POS processing is more accurate with a larger tagged training corpus. Since ML algorithms use training sets to understand patterns within the data, the more data available, the more instances of a pattern, or types of patterns, can be observed by the algorithms.

POS tagging is a critical piece to further processing of a corpus. Named entity recognition and Geoparsing, both described below, used POS in order to look for nouns that could be categorized based on their context in a sentence (Martinez 2010).

2.5.3 Named Entity Recognition

Named Entity Recognition (NER) is an optional step in the NLP pipeline that is used to sort word tokens into recognized categories (Sekine 2009). These categories include name, location, date, time, and currency. The NER process comes after the POS tagging stage and typically looks at words that have been identified as nouns and attempts to use the context of that word to assign a category. Early work (Kripke 1982) in the field of NER focused on the identification of proper names in general and then later expanded to differentiate between a name of a "person", "location", or "organization" (Thielen 1995).

NER, like other ML processes, looks for patterns with a word and the surrounding words to properly label word tokens with a category. On the word level, the processor looks for attributes like a words case, punctuation, or POS. The Word-level features table (Table 7) from shows these various attributes:

TABLE 7 WORD-LEVEL FEATURES

Features	Examples
Case	– Starts with a capital letter – Word is all uppercased – The word is mixed case (e.g., ProSys, eBay)
Punctuation	– Ends with period, has internal period (e.g., St., I.B.M.) – Internal apostrophe, hyphen or ampersand (e.g., O'Connor)
Digit	– Digit pattern – Cardinal and ordinal – Roman number – Word with digits (e.g., W3C, 3M)
Character	– Possessive mark, first person pronoun – Greek letters
Morphology	– Prefix, suffix, singular version, stem – Common ending
Part-of-speech	– proper name, verb, noun, foreign word
Function	– Alpha, non-alpha, n-gram – lowercase, uppercase version – pattern, summarized pattern – token length, phrase length

The NER process then looks at known lists of identified common words. These lists include common words that can be categorized by themselves, like days of the week or months of the year, or words that are often found in conjunction with others, like the word "associates" is frequently found in organization names (McDonald 1993).

Some words can appear as more than one category, like the word Washington, both a person's name and a location. The NER process attempts to use various patterns to identify which category the word should fall under, but that is not always possible, so the category with the higher probability of being correct is associated with the word or no category is assigned (Poibeau 2006).

<enamex type="location">Washington</enamex>

It is from the NER process that the standardized Enamex tags are added around word tokens to add the category data as an attribute. 'Enamex' tags are the standard tags added by most NLP systems during the NER process. This tag in digital file allows for changes in formatting when displayed or the potential for further processing. Using NER processing, words identified as locations can be sent to a geoparser for continued processing in order to assign coordinates and other geospatial data, typically added as additional attributes to the enamex tag (Sekine 2009).

2.5.4 Geoparsing

The geoparser is a software application that identifies place names in text through the use of NLP and disambiguates them using a gazetteer. A gazetteer, in this context, is a geographic dictionary containing place names and corresponding geographic data. In its simplest state a gazetteer is a list of place names and the associated latitude and longitude, but can contain information like alternate names, location type, administrative zone (Berman 2016).

The Edinburgh Geoparser[18] is a corpus linguistic tool developed by the Language Technology Group at the University of Edinburgh (Alex 2015). The Edinburgh Geoparser uses the Unlock[19] and Geonames[20] gazetteers, both of which are robust and peer reviewed databases. Edina's Unlock gazetteer is a U.K. based geoconsultancy resource with a focus on locations in the U.K., while the Geonames gazetteer is administered as an opensource archive of worldwide location information, primarily constructed from U.S. based government agency records. Using NLP and NER, the Edinburgh geoparser identifies place names within the processed text, and then searches each place name against those in the gazetteers. The NER process is used to identify words within a text that can be classified as 'locations.' The geoparsing process is used to assign words identified as 'locations' via the NER process with geographic coordinate data drawn from gazetteers. Each match is then annotated into

[18] http://groups.inf.ed.ac.uk/geoparser/documentation/v1.1/html/index.html
[19] https://edina.ac.uk/unlock/
[20] http://www.geonames.org/

the text as an ‘enamex’ XML tag. If a place name is not found in the gazetteers a blank ‘enamex’ tag is created so that a record exists for each place name. The sample of text below, from “The English Lakes”, shows the annotation of place names in these two instances:

> *This is a clean-looking ancient town in the south of the county of <enamex sw="w4910" name="Large admin unit">Cumberland</enamex>, lying near the junction of the Eamont and <enamex sw="w4964" long="-2.719930910627173" lat="54.6044092845645" type="ppl" gazref="unlock:11209204" name="Lowther" conf="2.6">Lowther</enamex>, and containing upwards of 6000 inhabitants*[21].

Each ‘enamex’ tag can contain any of the following depending on the availability of data in the gazetteers: the locations latitude and longitude (lat,long), place type (type), the source gazetteer (gazref), the place name or attributed descriptor (name), or the confidence of the geoparser in the match (conf). Once annotated by the geoparser, a text can then be visualized in applications like google maps using the xml tags, or the locations can be brought into GIS applications for spatial analysis (Gregory 2015).

2.5.5 Java and Web Platform

The case studies created for this research heavily use a custom Java web server back end to process and disseminate data to the web front end. Java is a widely used programming language that supports complex structures and allows for processing of data using internal algorithms and external application programming interfaces (APIs)[22]. APIs are usually pieces of code in the public domain which have been designed to perform a specific operation, or series of operations, and are available as open source products for integration into new applications. APIs allow for complex operations to be used without requiring new code to be written, saving time for other developers.

Web was chosen as the primary visualization and interaction interface for the prototype applications developed for this research. Web applications[23] are easily accessible by all types of users, only requiring a web browser. Web browsers allow for the visualization of relatively complex content without the intense graphical processing required by other dissemination methods. While other applications, like standalone PC or Mac programs can have much more powerful graphical rendering and computational power, web visualization is accessible on nearly all devices, an ideal

[21] Note: These enamex tags have had the NER type attribute removed, as all the enamex tags in the CLDW are of location category.

[22] https://developer.mozilla.org/en-US/docs/Web/API

[23] https://w3-lab.com/web-development/choose-web-application-over-desktop/

trait for Deep Mapping applications that aim to include various audiences in their use. The web is also a convenient platform because there is a plethora of open source datasets that can be drawn upon as layers in a Deep Map. As Deep Maps are inherently spatial products, mapping visualizations provide an ideal framework to build upon, the most useful of which is Google Maps.

The Google Maps API[24] is widely used in the following case studies. As one of the most used web mapping APIs it is a robust framework that provides significant functionality. The mapping piece of the API provides access to various map types, like roadmaps, satellite, and terrain, and has the functionality to insert additional map types from external sources, like the use of historic maps used in the following NLP and Close Reading case studies. Google also gives access to other spatial datasets from its massive collection through this API. One such dataset that is used in the applications developed for this research is the Google Street View dataset[25]. Google Street View is a collection of panoramic, 360 degree, images that have been captured along roads all over the world. Public users can also submit panoramic images to Google, so less accessible locations that are of public interest often have imagery that can be accessed through the Google API, including the tops of mountains or on the surface of lakes. The case studies below use an operation in the API that pulls the closest panorama to a given location to provide users with a view at or near the designated point.

2.6 Methodology

This PhD research focuses on exploring the development and use of Deep Mapping software applications, and falls within the interdisciplinary field of digital humanities, incorporating the disciplines of history, literature, computer science, and data science, as well as others. Given this interdisciplinary environment and the intention to develop software prototypes, a methodology was required that would enable the development of Deep Mapping application prototypes and provide a structure to answer the projects research questions through these prototypes. This project drew on the fields of information systems, HCI, interaction design, as well as past work in the digital humanities, to identify Research through Design as a viable methodology for exploring research concepts through the prototyping of Deep Mapping applications.

[24] https://developers.google.com/maps/documentation

[25] https://developers.google.com/maps/documentation/streetview/overview

Research through Design (RtD), or Research by Design, is a methodology that uses design as a method of investigation. RtD can be thought of as one of three forms of the possible permutations of the research and design relationship. Alain Findeli defined the three forms of design research as (Findeli 2008):

> *"Research for design" aims at helping, guiding and developing design practice. Those researches document the processes and concerns of professional designers and treat designers and their practice as the object of their study.*
>
> *"Research into design" is mainly found in universities and research centres contributing to a scientific discipline studying design. It documents objects, phenomena and history of design.*
>
> *"Research through design" is the closest to the actual design practice, recasting the design aspect of creation as research. Designer/researchers who use RtD actually create new products, experimenting with new materials, processes, etc.(Godin 2014)*

RtD focuses on the iterative process, using trial and error, to better understand complex issues in design problems. Projects that use RtD methodology usually have two similar constraints, "looseness" and "openness". Looseness in a project refers to the number of possible outcomes given the constraints set by the project. A problem set by a project that has various constraints that can be addressed in a number of ways has a looseness to it. Openness refers to the degree of outside influence on a project. An open project would allow for feedback and participation from sources outside the immediate stakeholders, designers or researchers, of the project. The openness of a project provides additional sources of input and allows for continued iteration of the design (Roggema 2017).

The concept of looseness applies well to Deep Mapping, given its non specific definitions. An example of this is the constraint that Deep Maps should allow for spatial navigation of data, which can be done in a number of ways, including via a map, a non-linear presentation, a 3D environment, or an abstract visualization. In this PhDs research, the looseness in Deep Mapping, allowed for the development of prototype Deep Maps that followed more conventional approaches, the interactive maps (Chapters 3-5), and less conventional approaches like the 3D Deep Map (Chapter 6).

Openness is also well suited for Deep Mapping projects in general, as open access and public feedback are closely linked with developing meaningful content associated with place, a key attribute of a Deep Map. This can be seen well in the Cork Deep Map project, where public testimonies and accounts have been incorporated into the cultural content displayed in the map. The prototypes developed in this PhD research also allowed for a level of openness within the projects, bringing diverse datasets like traveler's accounts, artists renderings, and personal testimonials into the Deep Mapping applications, as well as making those applications accessible to a wide audience for testing and feedback.

The RtD process allows for 'critical inquiry' through the act of design. It is a method by which new 'insights, knowledge, practices and products' can by created (Roggema 2017). While many disciplines use RtD as a valid methodology in research, they do not all follow the same structure in applying the methodology. This PhDs research is structured on the concepts defined by Ken Peffers at el., following the steps:

Problem Identification

Defining Objective of a Solution

Design and Development

Demonstration

Evaluation

Communication

The first step, problem identification, starts with identifying the desires of the stakeholders in order to define the problem and create viable constraints for the solution. The problem identified for a project should be defined well enough that a solutional can be created and evaluated based on its constraints, but still be loose and open, as discussed above, so that the problem allows for creativity and flexibility in the development of a solution. It is at this point that relevant data to the problem should be collected (Peffers 2007).

The second step, defining the objective of a solution, begins defining how a solution addressing the proposed problem will be constructed. Peffers defines three objectives of this phase:

> *"It will (1) provide a nominal process for the conduct of DS research, (2) build upon prior literature about DS in IS and reference disciplines, and (3) provide researchers with a mental model or template for a structure for research outputs. (Peffers 2007)"*

The definition of objective assists in the process of design and development and later supports the evaluation of the created solution.

The third step, design and development, focuses on the creation of a solution that meets the objectives proposed in step two, in order to address the problem introduced in step one. This step is experimental and exploratory, attempting to create a solution that works within the constraints of the problem and objectives, and is still achievable within the technological and research limitations of the time. A solution that requires technology that does not yet exist or requires a research scope that is not practical to implement, would not be a viable solution for the problems and objective and the wider research process.

The demonstration of the solution, step four, is intended to prove the idea works (Walls et al 1992). Demonstration of the solution allows it to be tested to help determine if it addresses the problem it was designed to address. This step helps to inform the iterative process. Results from demonstration inform if more design and development are necessary to create a viable solution for the problem.

The fifth step, evaluation, seeks to observe and measure how well the developed solution addressed the problem (Eekels et al 1991). This step involves comparing the objective of the solution to the observed result of its development in demonstration (Rossi et al 2003). Evaluation can take many forms depending on the nature of the problem. The evaluation step can also inform the iterative process and potentially lead back to step three if researchers determine the solution can be improved, depending on the scope and objective of the research.

The final step of the research through design process is the communication of results (Hevner et al 2004). Like with most research processes, the dissemination of resulting knowledge is a critical aspect to the impact of research. Communication of the process and results of the RtD process assist in the transfer of knowledge and can also provide insight and feedback that can inform continual iteration in the process. Like with demonstration and evaluation, the communication step can potentially lead back to step three, providing new insight that might improve the development of future iterations of the solution.

3 Case Study: Crosthwaite

3.1 Problem

To explore the concept of Deep Mapping as it applies to Digital Humanities and the Historic Lake District region, a limited scope project was needed, one that used a subset of the CLDW and was focused on a single topic. Using the full CLDW in the initial case study for this research would have focused more on the technical limitations of processing and visualizing such a large corpus and detracted from the design of a Deep Map application. Researchers on the Geospatial Innovations project requested an investigation into the creation of a Deep Map of the Picturesque Stations described by authors in a subset of the corpus. These Picturesque Stations, and the written entries of them, are significant because they affected how subsequent writers and tourists viewed the region. To further narrow the scope of this study, the researchers were also interested in the works of historic map maker Peter Crosthwaite, who illustrated maps of the Lake District region and depicted the Picturesque Stations in them. This case study focuses on the Derwent Water map and the eight Picturesque Stations around it.

The case study problem:

CSP1 - How can the Crosthwaite Maps be used to create a Deep Map?

CSP2 - How can a selection of texts from CLDW be incorporated into a Deep Map?

3.2 Background

The English Lake District is renowned for its literary and artistic history. It is famously a place of poetry and of painting. Crucially, though, it is also a place of maps. True, the region was among the last portions of England to be mapped in the first Ordnance Survey[26] in the mid 1700s. But the belatedness of the Lakeland's inclusion in the OS is significantly counterbalanced by its rich cartographic history. By the time the first OS maps of Cumberland appeared (in the 1860s), the Lake District was doubtlessly one of the most frequently and widely represented regions in Britain. Admittedly, many of these representations took the form of scenic pictures and verbal descriptions. Many, however, also took the form of maps, plans, and topographical diagrams. Among these

[26] http://www.ordnancesurvey.co.uk/about/history

maps, plans, and diagrams, few stand out as prominently as Peter Crosthwaite's 'Accurate Maps' of the Lakeland's eponymous lakes.

A native of the parish of Crosthwaite, near Keswick, Peter Crosthwaite (1735¬1808) had an exceptionally colorful career. As a master of a gunboat called the Otter, he spent his early adulthood protecting East India Company ships from pirates in the Bay of Bengal. Following his return to England in 1765, Crosthwaite opened a new chapter of his life working as a customs officer on the coast of Northumberland. Then in 1779 he returned to Keswick as a respectable man, married with two children (Fisher Crosthwaite 1778). The year after, he took a house in the Square and quickly set up shop as a purveyor of publications, amusements, and paraphernalia for tourists.

Of all Crosthwaite's business ventures, however, it is his pursuits as a cartographer that are of principal significance. In part, this is because of the peculiar mixed-media character of his maps (which we discuss below); in part, though, this is because his maps were the first designed specifically for Lakeland tourists. Whereas some of the early accounts of the Lakes region included cartographical plans, none of the early guide or tour books did. Even Thomas West's widely cited and celebrated Guide to the Lakes (1778) did not include a map until the publication of its third edition in 1784 the year after Crosthwaite began publishing his own 'Accurate Maps' of several of the region's key lakes. In this way, Crosthwaite's maps can be seen to complement other aspects of his entrepreneurial enterprises: they rushed to fill a gap in the region's developing tourist trade.

Crosthwaite's 'Accurate Maps' were produced between 1783 and 1794, and thereafter revised and reproduced until 1819. Crosthwaite completed his maps of Derwent Water, Windermere, Ullswater, and Pocklington's Island (in Derwent Water) in 1783. He enlarged this series in 1785 and 1788 with his maps of Bassenthwaite and Coniston Water. Finally, in 1794, he completed his series with a map of the western Lakes, which includes Buttermere, Crummock Water, and Loweswater. Meticulously surveyed and packed with detail, these maps are exquisite productions, and nearly all of them (save those of Windermere and Pocklington's Island) are drawn on the scale of 3 inches to the mile. All of them, moreover, are designed to be handheld; on average, they measure 18 inches by 8.5 inches.

These maps, after all, offered much more than information about the locations of places, roads, and routes. They were miniature compendia of topographical and picturesque detail. In addition to marking the principal landmarks, houses, and estates (as well as the names of their owners), Crosthwaite's maps featured vignettes of those landmarks, houses, and estates, as well as passages of descriptive poetry and prose. Beyond this, they also featured the locations of the viewing stations designated by Thomas West's Guide as those places where the Lakes scenery could be seen to its best advantage. Crosthwaite's creations could therefore be considered as prototype Deep Maps of their day, and so we considered them to be an excellent first case study in our own exploration of this so far ill-defined concept.

Julia Carlson has recently remarked on this curious 'convergence of cartographic and poetic innovation' (Carlson 2016 p.44). Crosthwaite's maps, explains Carlson, 'were a new form of print text intended to facilitate a new form of experience [...]. They encouraged physical exploration and demanded the interpretation of intersecting modes of measur[ing]' geometrical, pictorial, and literary (Carlson 2016 p.44). These composite creations, in short, combined different illustrative media not only to guide the tourist to key locations, but also to inform the tourist's feelings about those locations. These maps were tools for instructing not only the eye, but also the mind and the heart.

Thomas West's Guide to the Lakes (1778) was the primary focus of Crosthwaite's maps. The Guide to the Lakes described Thomas West's journeys through the Lake District region. West divided the locations visited in his journey into Picturesque Stations. At each of these stations, West described his thoughts, emotions and experience of the location. An excerpt from Station I in the area around Derwentwater can be seen below:

> *STATION I. COCKSHUT-HILL is remarkable for a general view, it is covered with a motly mixture of young wood, has an easy ascent to the top, and from it the lake appears in great beauty. On the floor of a spacious ampitheatre, of the most picturesque mountains imaginable, an elegant sheet of water is spread out before you, shining like a mirror, and transparent as chrystal; variegated with islands, that rise in the most pleasing forms above the watery plane, dressed in wood, or clothed with softest verdure, the water shining round them. The effects all around are amazingly great; but no words can describe the surprising pleasure of this scene, in a fine day, when the sun plays upon the bosom of the lake, and the surrounding mountains are illuminated by his refulgent rays, and their rocky broken summits reflected inverted by the chrystal surface of the water.*

Thomas West wrote about eight picturesque stations in the region around Derwentwater. As other writers traveled to the area, they used West's guide and wrote about the same locations in their writings. The CLDW contains eight texts from authors who wrote about the station in the Derwentwater region, including Thomas West.

3.3 Objective of the Solution

The objective was to develop a software application that brought together the historic Crosthwaite map, the Thomas West Picturesque Stations, descriptions of the Picturesque Stations by other authors in the CLDW, and other layers of data available through web APIs that would add to the experience of application, including maps, illustrations and panoramas. This software application was designed to incorporate these data sources in a spatial structure that met the requirements to make it a Deep Map.

Case study Objectives of the Crosthwaite Deep Map:

CSO1 - Incorporate the Crosthwaite Historic Map

CSO2 - Use texts from the CLDW

CSO3 - Have the layered data create a sense of place

CSO4 - Include multimedia relevant to the stations

CSO5 - Create links between the map, stations and text

CSO6 - Create a sense of exploration and discovery

3.4 Design and Development

This application[27] uses a Java web server to process the data on the back end, and an HTML front end to display the result to users. A Java web server was chosen because it is expandable and can handle a large amount of data, giving this application the potential to be expanded upon in the future. The application has an underlying SQL database which can be queried by the server and contains the metadata for the stations and text. The database also defines the sections of text, by sentence range or page range, which are associated with a station.

[27] https://github.com/areinhold17/deepmap

The main page (Figure 2) of this application is based on two map APIs, Google Maps and ArcGIS, modern and historic, centered on Derwent Water Lake. The Google Maps API is used as the modern-day view, and enables the use of panoramas, described below. The ArcGIS API was added to this application for the historic map view, which shows the Crosthwaite map projected onto the modern map. ArcGIS was needed for the historic view in order to keep a high resolution on the image of the Crosthwaite map, using Google's API would have required use of a degraded image. On both map views the picturesque locations are drawn as markers on the map. On the left side of the main page is a navigation panel. The navigation panel is broken into three parts. The first part, at the top, allows the user to switch between the historic map view and the modern map view. The second section of the navigation panel lists the picturesque station, showing the station number, latitude and longitude, and a map thumbnail of its location. The third section of the navigation panel is a list of the eight texts from the CLDW that are included in the application, showing their title, author and publication date.

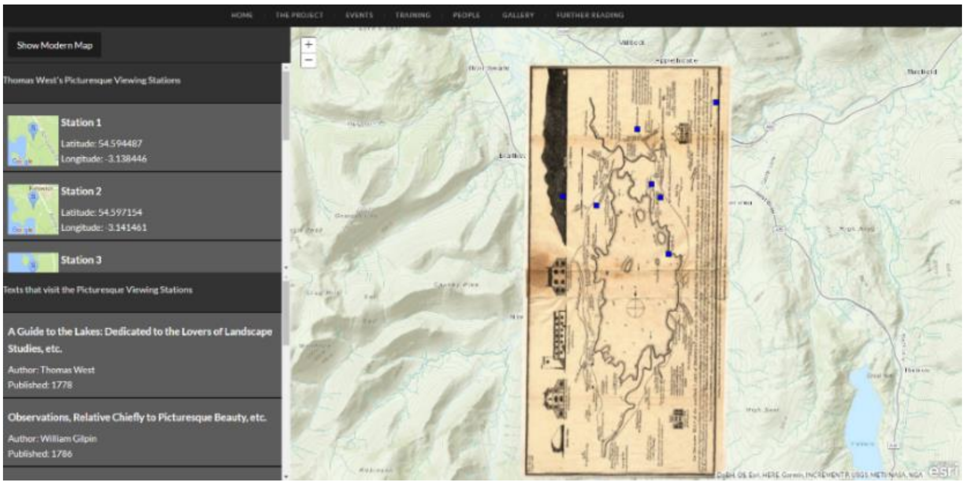

FIGURE 2 CROSTHWAITE DEEP MAP MAIN PAGE

By click on a station from the station list in the navigation panel, or by clicking on the corresponding marker on the map, the user is brought to the station view page (Figure 3). The station view is based on a Google Street view panorama of the station's location, or the closest panorama available to that location, based on its latitude and longitude. The panorama can be clicked and dragged to show the user the full 360 degree imagery of the site. Layered on top of the panorama in the station view are two panels that provide additional information about the station. The left panel shows excerpts from the eight CLDW texts that talk about the selected station (Figure 4).

FIGURE 3 CROSTHWAITE DEEP MAP: STATION VIEW

This panel is tabbed to allow the user to cycle through each author that described the station. Each tab shows the title of the text, the author, a link to the full text view, described below, for the selected text, and the section of that text that specifically talks about the station. An additional tab is present in this panel, which brings up a FoamTree (Figure 4). The FoamTree[28] is a Voronoi treemap visualization that shows the 50 most used words, excluding standard stop words, in all the text excerpts describing the selected station. Stop words are standard words that appear frequently, like A, Able, About, As, Is, It, The, etc., in most texts. The list of stop words used for this application were procured from a Google library[29]. The right panel in the station view shows illustrations drawn from the selected station.

[28] https://www.carrotsearch.com/foamtree/

[29] https://code.google.com/archive/p/stop-words/

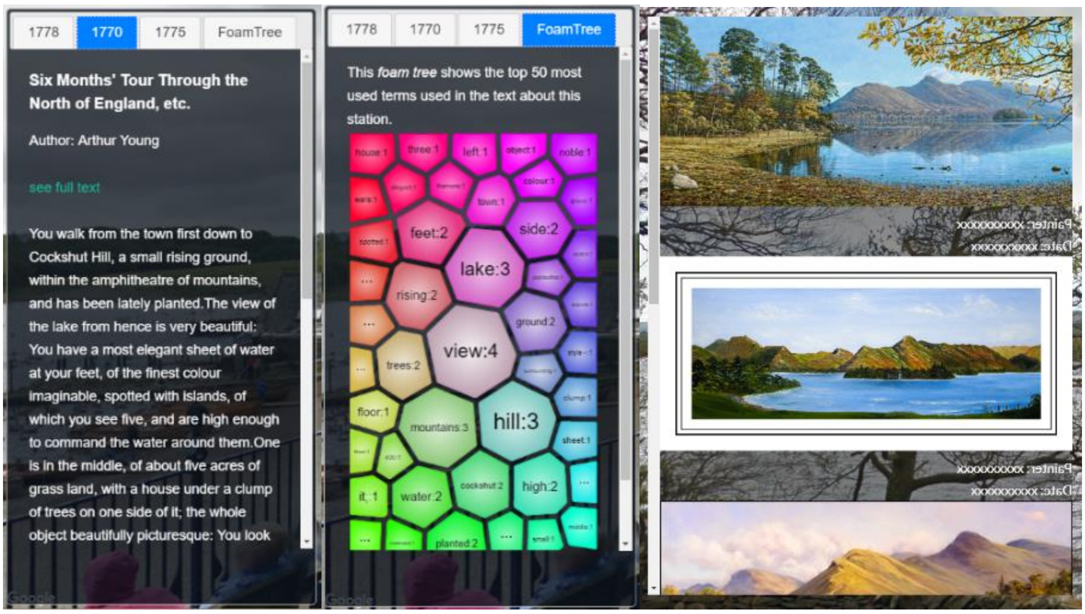

FIGURE 4 CROSTHWAITE DEEP MAP: STATION VIEW SUB ELEMENTS. (LEFT TO RIGHT) TEXT EXCERPTS, FOAM TREE, MULTIMEDIA PANEL

From the main page, by clicking on a text from the text list in the navigation panel, or by clicking on 'see full text' in the station view, the user is brought to the text view page. The text view allows the user to read the entire selected text, with additional interactive information displayed alongside. The text view page has a simple navigation in the top right allowing users to flip through the pages of the text, displaying one at a time. Each page shows an image of the original historic page on the right side and a list of the locations mentioned on that page on the left. If the page of text being viewed mentions a station, a link to that station is shown on the far right and would bring the user to the station view if selected. Clicking on location from the list on the right brings users to the main map view and focuses on the selected location. If illustrations are associated with a station mentioned on the page, those illustrations appear below the list of locations.

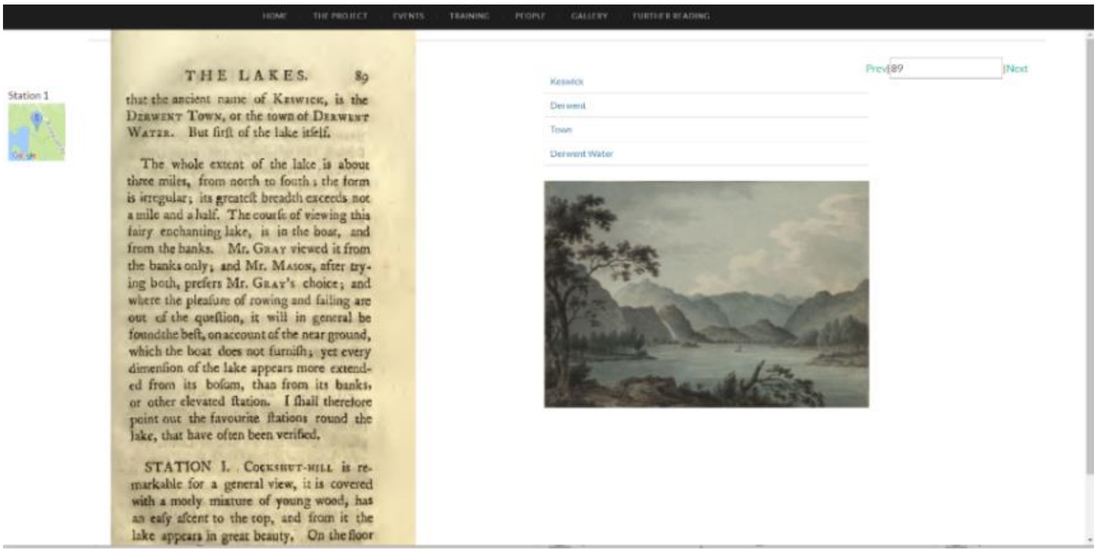

FIGURE 5 CROSTHWAITE DEEP MAP: TEXT VIEW

3.5 Demonstration and Audiences

After development of this prototype, the Crosthwaite Deep Map was presented at Conference on Spatial Information Theory (COSIT 2017) SPHINx workshop, for which a paper was published, and the Digital Past Conference (2018). The prototype was also presented at multiple seminars at Lancaster University. These events mostly included academics in Digital Humanities, History and Information Theory. The paper and presentations were positively received. This prototype was also presented at two Lancaster University Community Day events, in 2017 and 2018, which brought current research into public view. The presentation to the public was also positively received, with more significant interaction from participants who had an interest in the Lake District region.

This Deep Map did not have a specific audience targeted during the design of the application. The general public (UG3) ended up being the primary audience this map was effective for because it supports the exploration of the region that closely relates to tourism, especially those interested in literary history. This application is also effective for the other two user groups as a means of engaging with the data but doesn't provide substantial analytic interactions that would help make meaningful impacts to research activities or conservation efforts.

3.6 Evaluation

This section looks at the three categories of evaluation as defined in chapter 1.2. Each category has its own set of criteria and helps to provide insight into the effectiveness of this Deep Mapping implementation.

3.6.1 Design Evaluation

This case study sought to create a small scale prototype Deep Map that focused on the historic Crosthwaite maps. The goal of this prototype was to explore what a Deep Map might look like beyond the map, marker, and info box styles discussed in 2.4.2. Those previous projects, also prototypes themselves, were simple in design and lacked a greater depth to the interaction users had with the map and multimedia embedded in it. This case study's prototype application is more complex and layered than previous prototypes made by others. The ability to link back and forth between the map, the station view, and the text view allows for users to explore the Deep Map in multiple ways and focus on interactions that best engage them.

Problems:

CSP1 - How can the Crosthwaite Maps be used to create a Deep Map?

The Crosthwaite Deep Map qualifies as a Deep Map, as discussed below in 3.6.3. This case study sought to improve upon past work in the field. Previous prototypes typically focused on map interfaces with markers and info windows to display relevant data. This application took that strategy a few steps further by making the markers link to a new view which included various types of material and visualizations for the user rather than just a single info window with limited capacity. This strategy made each location on the map more than just a marker with some information, it made each location feel like a place with some substance. Users could see what each location looked like through the panoramas and artist's images and could read about what other travelers experienced while they were there. Linking from the map and the station view to the text view allows a strong connection to the literary works, providing context beyond the written word.

CSP2 - How can the CLDW be incorporated into a Deep Map?

In earlier iterations of the Geospatial Innovations Project, the CLDW was used in GIS in order to pursue various research questions as discussed in section 2.2. The GIS approach generally only used lists of locations and concordance matches derived from the CLDW texts. While this approach works well for some research questions, it can create a divide between the original texts and the resulting research results. This case study drew upon the CLDW and attempted to incorporate the relevant texts in a more holistic way, that would allow locations in the text to be mapped and still maintain their context within the original literary text. By giving users multiple methods of interacting with the texts, through the excerpts linked in the stations, to the complete version in the text view.

Objectives

CSO1 - Incorporate the Crosthwaite Historic Map

This task was relatively easy to complete in the building of the prototype. The mapping APIs allow for the importing of images as layers on the map through the use of KMZs[30]. In this application the ESRI mapping API[31] was used for visualizing the historic map because it allows for a higher memory limit, which means the historic map can be displayed at a higher resolution. If this prototype were to be expanded upon and put into distribution at a greater level, the use of a tile server would be appropriate to allow the historic map to be imported at the highest resolution to any mapping API, but given the limited scope of this project, that amount of work was not necessary.

CSO2 - Use texts from the CLDW

This objective was addressed early in the conception of this case study. The researchers interested in Crosthwaite's map provided a subset of the CLDW that discussed the picturesque station. Some additional work was required to determine if any other texts in the CLDW discussed the same locations, so that each location would have at least a few excerpts from different writers. As all of the CLDW was processed through a geoparser, all the locations were georeferenced and could easily be linked to the map. New tags had to be added into the chosen texts for this case study in order to denote the start and end of the passages that spoke about each picturesque location.

CSO3 - Have the layered data create a sense of place

This object was the focus of the station view. While it is difficult in a virtual environment to replace the experience of physically being somewhere, the goal of the station view was to provide users with enough information and visualization that they could feel like they had some understanding of the place. The panorama, text excerpts and artistic imagery all contribute to building a sense of place for interested users. The map and complete texts provide a wider context to the stations so that they do not become a place detached from space. Without the map and texts, it is possible that a user could experience the station view and identify with the location as a place, but might not understand the wider context of that place and where it is in relation to everything else, hence a detachment from space.

CSO4 - Include multimedia relevant to the stations

[30] A KMZ is a common file format for embedding images into web maps

[31] https://developers.arcgis.com/

The use of the panoramas was the primary multimedia incorporated into this prototype, apart from the texts and map. The panoramas provide a visual record of each location and context for the text excerpts discussing those places. While the panoramas and accompanying text would likely have been enough to create a sense of place and solidify the prototype as a Deep Map, the addition of artistic images added an additional layer of meaningful expression that might induce greater connection between the location and the users.

CSO5 - Create links between the map, stations and text

CSO6 - Create a sense of exploration and discovery

CSO 5 and 6 are addressed together as they overlap. CSO5 is the means by which CSO6 is addressed in this prototype. The concept of exploration and discovery is a key element of Deep Mapping. While providing a map with interactive markers is enough to meet this requirement at its most basic level, this case study sought to provide users with more avenues of exploration through the application. The Crosthwaite Deep Map allows users to discover stations on the map or by reading through the texts. Users can find the texts through the linked excerpts in the stations or the complete text links on the map. The map can be accessed through each station via the location links on each page of the text in the text view. This interconnected navigation was designed to create a strong link between the map, stations, and text, and provide users with the freedom to explore the application on their own terms.

3.6.2 Research Evaluation

This section of the evaluation looks at how this prototype helps to address overall research questions of this thesis.

RQ1 - How can Deep Mapping, in combination with spatial analysis, natural language processing, and corpus linguistic techniques, be used to inform the research process?

In this case study, the prototype was not intentionally designed as a tool to help inform the research process, but rather the design of the Deep Map prototype helped in the research process in this PhD research. Designing this prototype allowed for a controlled exploration of the practical implementation of Deep Mapping concepts.

RQ2 - What elements are required to best present historic data to various user groups in a way that makes it accessible and understandable? What visualization techniques best engage these different user groups?

This application, being the first undertaken as part of this PhD's research, experiments with a variety of methods to present historic data to various user groups. The use of maps, text, standard and 360 imagery and the use of the foam tree, were all incorporated into the application to understand their impact on dissemination data to the three user groups of interest. Feedback during public demonstrations of this application showed that the general public had interest in the map and station view sections of the application, but limited interest in the original texts unless the individual had some previous knowledge or interest in the literary works included. Sectioning out excerpts of the text based on the picturesque locations made it easier for the average user to understand the historic data in the application.

The foam tree was best received by academics during these demos, as it shows word frequency, commonly used by researchers, but did little for the general public. The panoramas were the most engaging for all the audiences who used the application because it is visually appealing and allows exploration of the imagery through its interaction.

3.6.3 Data Evaluation

This section looks at how well the prototype Deep Map fits within the definitions of Deep Mapping and related values and methods from the Spatial and Digital Humanities. An evaluation rubric was created for this thesis so that these elements could be assessed and compared. The rubric also includes a breakdown of this thesis's research questions in order to provide a base for comparison. The answers to each criterion in the rubric were derived from the evaluation survey, in person demonstrations, and seminar and conference presentations.

The categories and criteria in the evaluation rubric are discussed in chapter 1.2. Each criterion can be answered with a 'Yes', 'No' or 'Maybe'. 'Maybe' is used when the prototype partially addresses the criteria but does not fully fulfil the expected outcome of that criteria. For each criterion there is a comment field summarizing how this prototype addressed that criteria.

TABLE 8 CROSTHWAITE DEEP MAP EVALUATION RUBRIC

	Yes	**No**	**Maybe**	**Comment**
Digital Humanities Criteria				
Iterative Process	x			This Deep Map was developed through several phases and was improved upon using feedback from researchers involved in the project.

Experimental	x			As part of a PhD research project exploring the potential of Deep Mapping applications, this Deep Map is considerably experimental. Additionally with the lack of practice implementations of Deep Mapping applications, new applications like this one are experimental.
Collaborative	x			A multidisciplinary team of researchers provided feedback on this application to inform the iterative process.
Multimodal	x			Work on this application was published via traditional print. The application itself is presentable via the web and contains various multimedia elements including text and images.
Open and accessible	x			Built in as a web application this Deep Map is accessible to all audiences. The code for this application is available as open source code on GitHub.
Spatial Humanities Criteria				
Relationship between space and time			x	While time is represented by the text excerpts linked to each station by navigating through the tabs, this is not a primary interaction and could use more emphasis to really create a relationship between space and time.
Place built through multi-layered data	x			Each location in this Deep Map provides imagery and textual accounts through time to convey meaning to the audience.
Deep Mapping Criteria				
Spatial Exploration as a primary form of interaction	x			This application is explorable via both an interactive map and through the navigation of 3D panoramas.

Temporal context as a primary form of discovery		x		The temporal element of this application is not a primary means of navigation through the data.
Data from verifiable source for legitimacy	x			The primary data for this application is from the CLDW which is a verified historic literary source.
Layers fill targeted area	x			The target area of this Deep Map was limited to the scope of the Crosthwaite map, and the picturesque stations were spread around the region.
Accessible to broad audience	x			As a web application this prototype was accessible to a wide audience. The application itself has a standard interface design and its functionality is accessible to an audience with standard web browsing experience.
Research Question Criteria				
Inform Research Process	x			This application provides the potential for new research processes to be discovered by the linking of various data sources in a spatial context.
Accessible and understandable				This prototype was presented in multiple instances to all three user groups. All audiences were able to access and understand the application.
Researchers and Scholars	x			
Site Authorities	x			
General Public	x			While the public could understand the functional elements of this application, not all were familiar enough with the historical and literary content to fully understand the application's purpose.
Engaging				Feedback from users at presentations and demos showed

				this application was sufficiently able to engage audiences.
Researchers and Scholars	x			
Site Authorities	x			
General Public	x			
Repeatable			x	The Deep Mapping application relies on the concept of picturesque stations to create explorable locations and segment the related texts. In order for this process to be repeatable with other data sources there would need to be a focus on location classification(s) (ex National Parks, Churches or Castles).
Expandable	x			The Crosthwaite picturesque maps were made for many of the lakes within the Lake District. While this Deep Map focused on a single Crosthwaite map, this application could easily be expanded to include the others.
Total	18	1	2	

This prototype lacked a meaningful engagement with the temporal aspects of the data. While the Station View divided the excerpts by the years they were published, this was not a significant feature and was a minor interaction in the navigation through the data. The lack of temporal interaction does not disqualify this prototype from being a Deep Map as the majority of the available definitions of Deep Mapping do not describe a temporal requirement.

The other criteria that this prototype did not entirely fulfil was related to repeatability. With the concentration on picturesque stations, this application requires a similar location-based focus. Location based elements that can be categorized, like churches, or heritage site, would be sufficient for integrating other datasets into this prototypes framework, but the categories used to display locations is defined during the development stage of this prototype, and doesn't allow for users to define them, like they can with the queries performed in the Analytic Deep Map described in chapter 4.

3.7 Conclusion

This case study has allowed us to explore Deep Mapping in a historical context. As our overarching thesis is heavily centred on the literary texts of the Lake District, our Crosthwaite map has demonstrated how the text and spatial elements can be brought together. Understandably, the creation of a historic Deep Map is complex and will require thought as to which features will bring the most value to the application. The representation of time, and the historic nature of the content, is another key consideration that emerged from our study and will need further thought. As highlighted here, this impacts both on user interface issues (e.g. visual representation of the original pages of the books) as well as side-by-side presentation of materials (such as original drawings, paintings and modern web-based panoramas). This application has revealed that our future Deep Mapping applications will need to consider how to maintain the historic context of the data using different visualization technologies, and how to represent ambiguity in the spatial features found in the data. Our prototype provides a model for exploration, and focus on linking historic content to the visual representation of place in a Deep Mapping application. This prototype fits within the 'person' category of the conceptual model proposed by Ridge et al, as described in chapter 2.4.2. The limited scope of this prototype and the focus on the literature associated with the picturesque locations makes this Deep Map closely related to a spatial narrative, while still allowing users to freely explore the application on their own terms and without curation of the path they take through the data.

4 Case Study: Analytic Deep Map

4.1 Problem

After the Crosthwaite Deep Map case study, to further explore Deep Mapping concepts, there was a need to expand the scope of data used to create a Deep Mapping application. In order to address this need, the second case study undertaken for this research was the creation of a Deep Map of the entire CLDW. Given the size of corpus, the diversity of content, and the variability in the needs of the target audiences described in chapter 1.5, two solutions were designed. The first was the Analytic Deep Map, focusing on analytical capabilities and the visualization of large queries against the CLDW dataset. The second, described in chapter 5, the Close Reading Deep Map, focused on the user's interaction with the text.

The case study problem:

> **CSP1** - How can corpus linguistics methods be a primary method of interaction in a Deep Map?
>
> **CSP2** - How can a Deep Map be designed to be an analytic tool using the full CLDW?

4.2 Background

Research in the Spatial Humanities has often used tools like GIS, to analyze the spatial relationships of quantitative data, or corpus linguistic tools, to analyze larger sets of textual data. In recent years these methods have been used together to look at texts in a geographic context. This combination of methods is called Geographic Text Analysis (Gregory et al 2015, Murrieta-Flores et al 2015). Geographic Text Analysis (GTA) makes use of NLP and geoparsing tools to identify and geolocate locations within text, as discussed in section 2.5. These locations, in conjunction with corpus linguistic techniques, like concordance matching, are used to map the texts. A study on nineteenth-century mortality (Porter et al 2015) used these methods to map diseases within their corpus of Registrar-Generals Reports. Concordance matches were created for each disease the team was targeting, such as Bronchitis, Influenza and Scarlet Fever. Locations returned within the concordance matches were tabulated and brought into a GIS environment in order to map each disease. Using the dates of the Registrar-Generals Reports, the spread of each disease could be visualized over time and standard GIS practices could be used to analyze the data.

This method of Geographical Text Analysis, while an effective tool for Spatial Humanities Research, can be time consuming, with processing of a corpus, determining what search terms are of interest, preparation of concordance matches, and integration into GIS all require time and expertise to achieve. This limitation of the GTA process is what the Analytic Deep Map is aimed at addressing, allowing the quick prototyping of research projects, so that time spent on the GTA process is not wasted pursuing unusable or ineffective results.

4.3 Objective of Solution

The Analytic Deep Map was focused on addressing one of the main research questions, RQ1, looking at what new research processes are supported by the affordances of interactive Deep Map methods that combine spatial analysis, NLP and corpus linguistics techniques. The objective of this application prototype was to create a platform that would allow NLP and corpus linguistic analysis to be performed in a spatial context. The focus of this Deep Map was centered on the ability to query the CLDW using concordance matches around user defined search terms, and visualize locations returned in the concordance matches. The aim of this Deep Map is to provide researchers the ability to prototype their research leading to the GTA process. By allowing researchers to quickly conduct searches of the corpus using concordances, they can experiment with search terms in order to identify what terms have spatial relationships worth pursuing in the more formalized GTA process. The discussion of this prototypes design uses the search terms of 'waterfall' and 'tourism' as examples, as these words have potential interest to researchers and are likely to be mentioned in relation to named locations, leading to search results that demonstrate the use of this tool.

Case study objectives:

- **CSO1** - Use the full CLDW
- **CSO2** - Use Corpus Linguistics concepts for data interaction
- **CSO3** - Provide layered data for each location to create a sense of place
- **CSO4** - Use time as a key function of application interaction

4.4 Design and Development

This application[32], like the Crosthwaite Deep Map, uses a Java web server to process and manage data on the backend and make queries to an underlying SQL database. The front end for this application used HTML and JavaScript to call data from the server and visualize it. Given the dataset for this case study was ten times the

[32] https://github.com/areinhold17/deepmap

size of the Crosthwaite study, and that the backend would perform a significant amount of processing for user searches, the expandable nature of the Java web server was necessary.

The main page of the Analytic Deep Map is built around a Google Map view, centered on the Lake District (Figure 6). This map view uses an API from the National Library of Scotland[33] to display historic Ordinance Survey maps of the United Kingdom, rather than the standard modern Google Map imagery, though the user still has the option to switch between these two visualizations and a modern satellite view. At the top of the main page are two horizontal panels, the search panel and the timeline panel. On the left of the main page is a results panel, populated by matches from the search function.

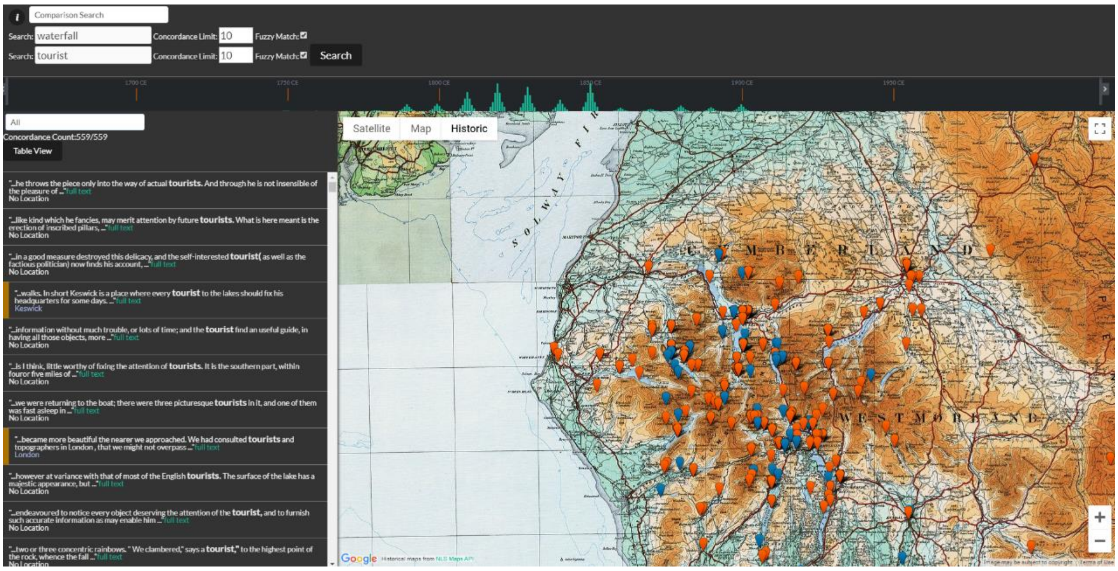

FIGURE 6 MAIN VIEW OF THE ANALYTIC DEEP MAP

Search Panel

The search panel starts with a drop-down menu in the top left that allows the user to select from five different search types. These types include a Simple Search, Advanced Search, Comparison Search, Advanced Comparison Search, and Location Search.

FIGURE 7 ANALYTIC DEEP MAP: SIMPLE SEARCH VIEW

Simple Search

The Simple Search (Figure 7) has four input fields, the search term, the concordance limit, a toggle for 'Fuzzy Match', and text selection.

[33] https://www.nls.uk/

The search term can be any word or phrase the user is interested in searching.

Table 9 Example of Fuzzy Concordance Matches

Search Term: Lake
Concordance Limit: 10
Searching Excerpt from Wordsworth's "Directions and Information for the Tourist Directions and Information for the Tourist (1835)": *In preparing this Manual, it was the Author's principal wish to furnish a Guide or Companion for the Minds of Persons of taste, and feeling for Landscape, who might be inclined to explore the District of the* **Lakes** *with that degree of attention to which its* beauty *may fairly lay claim. For the more sure attainment, however, of this primary object, he will begin by undertaking the humble and tedious task of supplying the Tourist with directions how to approach the several scenes in their best, or most convenient, order. But first, supposing the approach to be made from the south, and through <enamex sw="w695" name="Large admin unit">Yorkshire</enamex>, there are certain interesting spots which may be confidently recommended to his notice, if time can be spared before entering upon the* **Lake** *District; and the route may be changed before returning*.
The text in orange and the text in green are the two concordance matches returned by searching the term 'Lake' within this paragraph. Each match includes 10 words on each side of the search term as defined by the concordance limit. These matches of 21 words each are examples of what is returned by a search in the Analytic Deep Map.

The concordance limit is the number of words on each side of the search term that will be returned for each match of the search term in the CLDW, but default 10. The concordance limit is represented by the green and blue words in Table 9 Example of Fuzzy Concordance Matches. The contents of XML tags, like the enamex, are not considered as part of the concordance limit.

The 'Fuzzy Match' toggle, by default on, will determine the exactness of the search. If off, the search will only return an exact match to the search term. In Table 9, the orange match would not be returned for the search term 'Lake' because 'Lakes' is not an exact match. If on, the search will return any match that is an exact match to the search term or a match that contains the search term. Both matches in Table 9 would return if Fuzzy Match is on, since the search term 'Lake' exactly matches the green match and the orange match of 'Lakes' contains the search term 'Lake'.

The text selection button opens a window that allows the user to select a subset of texts from the CLDW rather than the whole corpus.

FIGURE 8 ANALYTIC DEEP MAP: ADVANCED SEARCH

Advanced Search

The Advanced Search (Figure 8) has all the same fields as the Simple Search but adds an additional search term.

The Second Search Term adds a constraint to the simple search, only returning matches of the first search term that contain the second search term within the concordance limit. In Table 9, with the Second Search term of 'beauty', only the orange match would be returned.

FIGURE 9 ANALYTIC DEEP MAP: COMPARISON SEARCH

Comparison Search

The Comparison Search (Figure 9) is two Simple Searches occurring side by side. The results from the two searches are displayed together in the interface but in two different colors, allowing for a visual comparison of the results.

FIGURE 10 ANALYTIC DEEP MAP: ADVANCED COMPARISON SEARCH

Advanced Comparison Search

The Advanced Comparison Search (Figure 10) is two Advanced Searches side by side, allowing for a comparison to be made between two constrained searches. The results of this search are also displayed in two different colors in the interface.

FIGURE 11ANALYTIC DEEP MAP: LOCATION SEARCH

Location Search

The Location Search (Figure 11) provides a different type of search in comparison to the other. The first four searches use search terms to return concordance matches found in the CLDW, while the Location Search uses a geographic location as the search term and returns a concordance match around any tagged place name (enamex) that contains a latitude and longitude coordinate in the search zone (Figure 12). The Location Search has four fields, including the latitude and longitude, radius and concordance limit. The Latitude and Longitude fields allow the user to manually enter a known coordinate to be used as the center point of the search. The radius is the number of kilometers around the center point that will be searched, creating a circular search zone. There is also the option Pick Point button, which allows the user to pick a point on the map just by clicking, rather than manually entering in the latitude and longitude.

This search would return any enamex tag in the processed CLDW that contained a coordinate within the red circle. Ex “<enamex sw="w5322" long="-3.125370693283439" lat="54.60160037173823" type="ppl" gazref="unlock:10824" name="Keswick" conf="2.6">Keswick</enamex>”

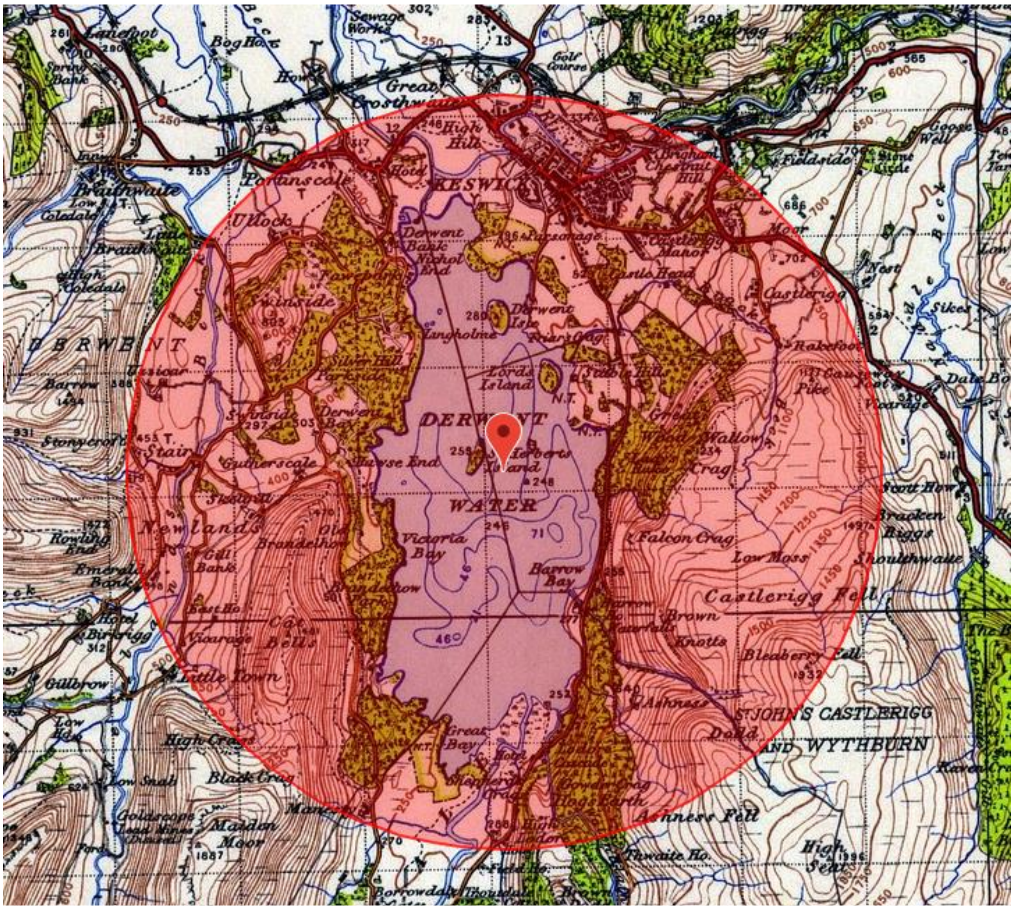

FIGURE 12 ANALYTIC DEEP MAP: LOCATION SEARCH RESULT

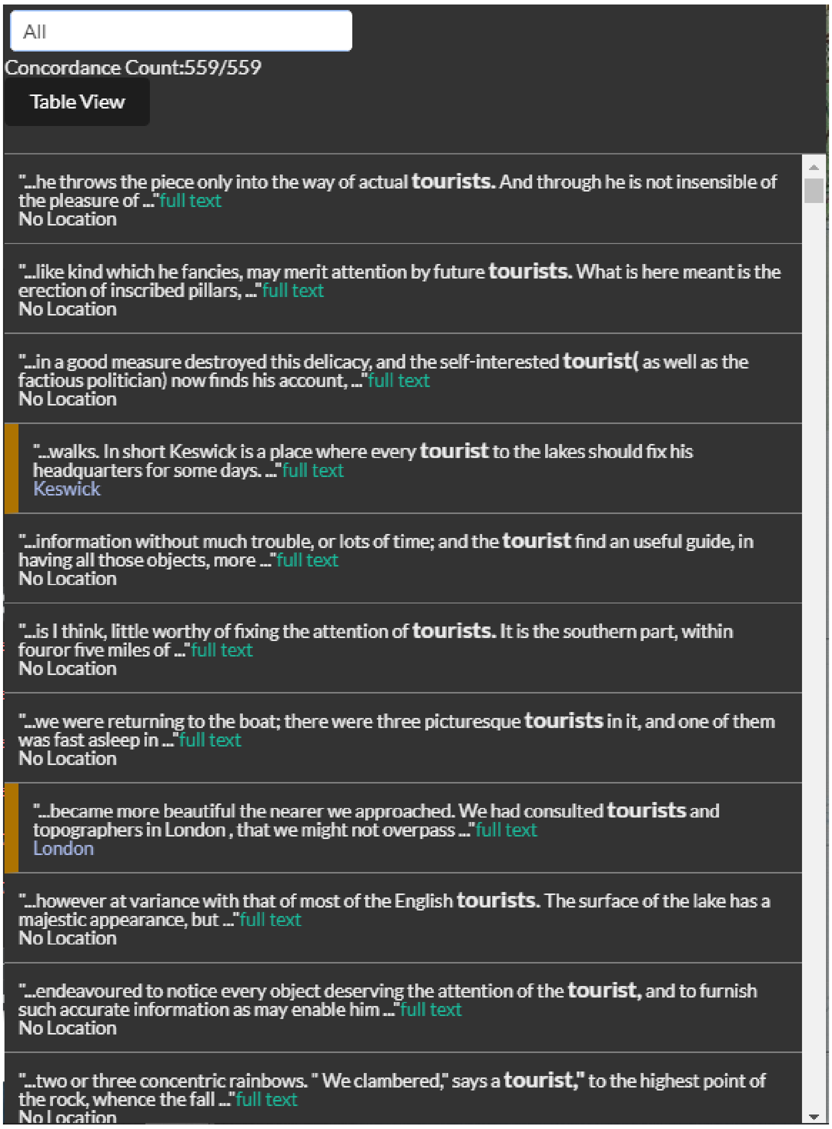

FIGURE 13 ANALYTIC DEEP MAP: SEARCH RESULTS PANEL

Results Panel

Each search type sends a request to the Java server, which then processes the request against all the selected CLDW texts and returns all the matches back to the front end. All the returned matches are displayed on the left results panel (Figure 13). The results panel header has a dropdown at the top allowing the filtering of the matches between those that have locations within the concordance and those that don't. Below the filter is the Concordance Count, showing the number of matches being displayed, if filtered, versus the number of total matches. Below the count is a button to bring up the table view of the returned matches. After the header is a list of all the matches being displayed. Each match displayed the concordance that was found, with the search term in bold. After the concordance is a link to the full text that that match came from. This is followed by a list of all the place names identified in the concordance, if any. Each place name is clickable and will center the map on that location if selected. If a comparison search was performed, the left side of each match entry with at least one location will have a colored bar, blue or orange, which corresponds to the search it was returned by and matches the coloring of the markers drawn on the map.

When a search returns matches, in addition to populating the results panel, all place names within the concordance, that have latitude and longitude information within the enamex tag, are drawn as markers on the map. If the comparison search was used, the results from the two searches are drawn in two different colors, blue and orange. Clicking on a marker will bring up an info box showing the concordance that

FIGURE 14 ANALYTIC DEEP MAP: MAP MARKER AND INFO WINDOW

corresponds to the selected location (Figure 14). If more than one match mentions the same location, clicking the marker for that location will change the visualization, drawing all the instances of that location as markers in a circle around the original location, 'spidering' out from the clicked marker (Figure 15).

FIGURE 15 ANALYTIC DEEP MAP: OVERLAPPING MARKERS - SPIDER VISUAL

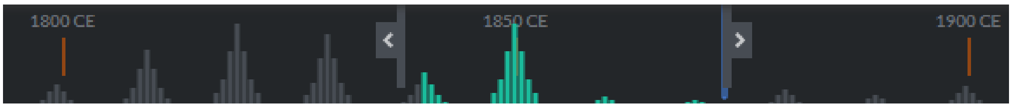

FIGURE 16 ANALYTIC DEEP MAP: TIMELINE

Timeline

Search results also populate the timeline in the top panel. The timeline ranges from 1600 to 2000, ensuring coverage of all texts in the CLDW. A histogram of the search results, based on the date of the text each match is drawn from, is layered on the timeline to show the relative number of matches for each period. The timeline can be adjusted with draggable sliders, as seen in Figure 16, which constrain the active time period. The constrained timeline filters the search results displayed in the result panel and on the map, showing only matches that are drawn from texts written within the selected time range.

Table View

The Table View page is available from top navigation on the left Results panel. The Table View displays the search results in a more traditional table. The left column of the table shows X number of words to the left of the search term, where X is the concordance limit. The center column shows the portion matched to the search term, either exact or fuzzy depending on the search settings, for example the search term of 'waterfalls,' with Fuzzy match on, in Table 10 returns 'waterfall' and 'waterfalls.' The right column shows X number of words to the right of the search term, where X is again the concordance limit. The Table View is intended to display the returned search results like other corpus linguistic tools, like wmatrix, so that results can be more easily compared across applications (Rayson 2008).

TABLE 10 ANALYTIC DEEP MAP: CONCORDANCE TABLE VIEW

dreadful heights the eagles build their nests: A variety of	waterfalls	are seen pouring from their summits, and tumbling in vast
the rugged cliffs, the steeps, the hanging woods, and foaming	waterfalls;	while the grand pencil of Poussin should crown the whole
the hithermost shore. at distance heard the murmur of many	waterfalls	not audible in the day-time. wished for the Moon, but
his idol gods and died. This alludes to the great	waterfall	of Lawdoor. Ah! where is he that swept the sounding
a Shower of Rain& offered to shew me a curious	waterfall	near his house. Returned by the Borrodale Road& after
did not look near so fine as at Sea. The	waterfalls	sounded dreadfull. It grew dark, windy& they missing me, I

The Table View page also has a tabbed view on the right with additional information about the search results. The first tab in this view is a frequency table, showing the frequency of every word returned as part of the search result matches. The second tab shows a FoamTree visualization which displays the 50 most used words in the returned search results.

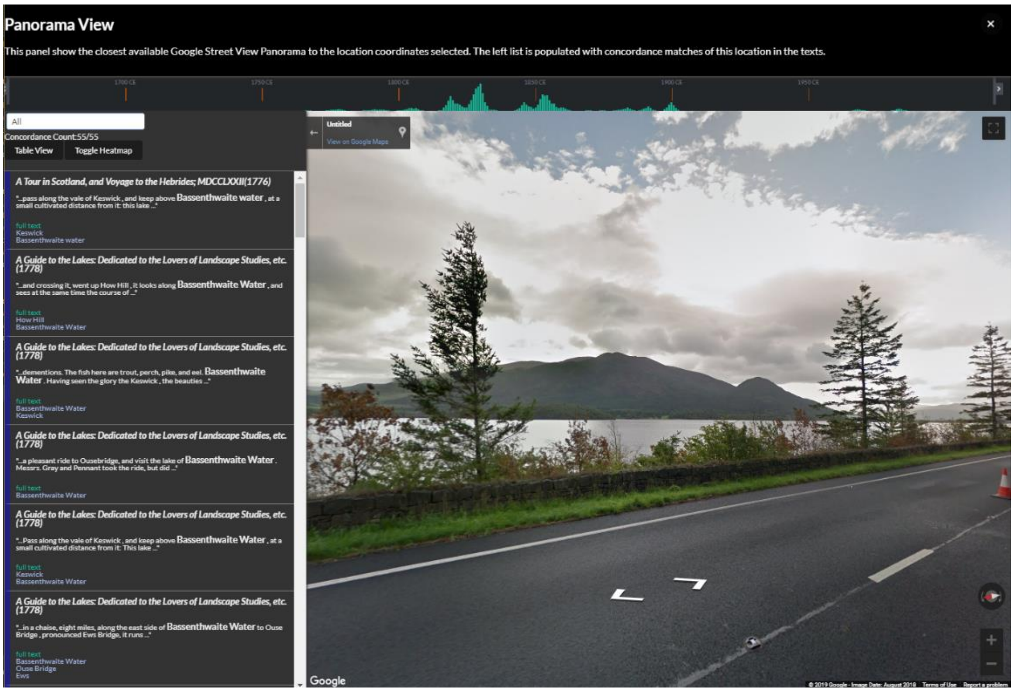

FIGURE 17 ANALYTIC DEEP MAP: MARKER VIEW

Marker View

The Marker View is a feature added in the second round of development of the NLP map (Figure 17). This view is accessible from the info box popup that appears when clicking on a mark. Clicking on a link in the info box bring the user to the Marker View. This view replaces the map with a 360 panorama of the selected location, using Google Street View. Upon entering this view the application does a secondary search, using the location selected as the search term, and returns concordance matches to a results panel on the left. This feature allows the user to see how a specific location is being talked about throughout the texts, while viewing imagery of the site.

4.5 Demonstration for initial formative feedback

This prototype was demonstrated at seminars held at the University of Lancaster, and public engagement events. In one of the first demonstrations of the concordance prototype to the Geospatial Innovations team at Lancaster, one response questioned whether it qualified as a Deep Map. In its original form, this application did not include a Marker View for locations and so while there was a significant amount of data being presented to the users, it was not presented in a manner that built a sense of place. In order to address this issue, the station view concept from the Crosthwaite map was added to the concordance prototype to help establish place. By providing a visual record of a location, alongside excerpts of author's mentions of that location, a greater connection could be created between the users and the locations they explored.

4.6 Evaluation

This section looks at the three categories of evaluation as defined in chapter 1.2. Each category has its own set of criteria and helps to provide insight into the effectiveness of this Deep Mapping implementation.

4.6.1 Design Evaluation

This section looks at how the problem and objective stated in 4.1 and 4.2 have been addressed by the design and development of the Analytic Deep Map.

Problem Evaluation

CSP1 - How can corpus linguistics methods be a primary method of interaction in a Deep Map?

This prototype's goal was to use corpus linguistic methods to allow for the exploration of the data. This was solved with the use of concordance for the querying of data to display on the map. Concordance matches proved an excellent solution to query the data because it allows for the searching of any terms of interest to users and still provides enough context around the term based on the concordance limit. The words returned around the term made the search results mappable when they contained geolocated place names.

CSP2 - How can a Deep Map be designed to be an analytic tool using the full CLDW?

Researchers are often interested in the significance of terms within a corpus. This application allows for the visualization of queries against the CLDW. Visualizing the coverage of a single word or comparing two words across the region enables researchers to see one aspect of how a word might be significant. GIS work from the Geospatial Innovations projects looked at the significance of the words 'beautiful' vs 'sublime' and how they were spread through the lake district region (Donaldson 2017). That work showed that there were differences in where those words were used. The type of analysis conducted in that study could be easily undertaken using the Analytic Deep Map with significantly less time invested.

Objectives Evaluation

CSO1 - Use the full CLDW
This application makes full use of the CLDW. All texts within the corpus are searchable via the applications search features. This application uses a directory folder to store the raw texts from the corpus and can be easily expandable by adding additional texts to the folder. The texts require the geoparsing markup undertaken earlier in the project in order to geolocation place names returned in the search results.

CSO2 - Use Corpus Linguistics concepts for data interaction
As covered in the first section of the problem evaluation, this application focuses on using concordance to conduct queries into the CLDW. This application also uses concordance and word frequency to display the results of searches.

CSO3 - Provide layered data for each location to create a sense of place
Like the Crosthwaite Deep Map, this application uses the same type of interaction with the markers on the map to bring up a location view with a panorama. This section shows all the matches within the corpus based on the location, providing the writers accounts of that location. Those accounts along with the panoramic imagery builds a sense of place for users interested in the location.

CSO4 - Use time as a key function of application interaction
The timeline element in the prototype allows for the temporal nature of the data to be a key component of searches. The timeline restricts the displayed results based on the desired time range. This functionality was key to making time a critical interaction in the application. Given the CLDW spans over 200 years, the time period that each text is from is important to understanding its context. The timeline element also provides an additional utility for researchers querying the data.

4.6.2 Research Evaluation

This section of the evaluation looks at how this prototype helps to address overall research questions of this thesis.

RQ1 - How can Deep Mapping, in combination with spatial analysis, natural language processing, and corpus linguistic techniques, be used to inform the research process?

The Analytics Deep Map was designed to be a tool to help research query the CLDW and assist in informing the research process. The second question in the problem evaluation section above touches upon this research question. The ‘beautiful’ vs ‘sublime’ study (Donaldson 2017) shows that the Analytic Deep Map can be used to effectively pursue research questions based on corpus linguistics methods.

RQ2 - What elements are required to best present historic data to various user groups in a way that makes it accessible and understandable? What visualization techniques best engage these different user groups?

This map was heavily targeted at creating a tool for academic researchers and can be confusing for other user groups to understand the corpus linguistic components. The use of two marker colors, rather than changing marker styles or shapes, for comparison searches was a useful strategy to display the different searches while keeping a clean presentation. Like the Crosthwaite Deep Map, the panoramic images were an engaging aspect to the application and promoted investigation and exploration of locations returned in searches.

4.6.3 Data Evaluation

This section looks at how well the prototype Deep Map fits within the definitions of Deep Mapping and related values and methods from the Spatial and Digital Humanities. An evaluation rubric was created for this thesis so that these elements could be assessed and compared. The rubric also includes a breakdown of this thesis's research questions in order to provide a base for comparison. The answers to each criterion in the rubric were derived from the evaluation survey, in person demonstrations, and seminar and conference presentations.

The categories and criteria in the evaluation rubric are discussed in chapter 1.2. Each criterion can be answered with a 'Yes', 'No' or 'Maybe'. 'Maybe' is used when the prototype partially addresses the criteria but does not fully fulfil the expected outcome of that criteria. For each criterion there is a comment field summarizing how this prototype addressed that criteria.

TABLE 11 ANALYTIC DEEP MAP EVALUATION RUBRIC

	Yes	No	Maybe	Comment
Digital Humanities Criteria				
Iterative Process	x			This Deep Map was developed through several phases and was improved upon using feedback from researchers involved in the project.
Experimental	x			As part of a PhD research project exploring the potential of Deep Mapping applications, this Deep Map is considerably experimental. Additionally with the lack of

				practice implementations of Deep Mapping applications, new applications like this one are experimental.
Collaborative	x			A multidisciplinary team of researchers provided feedback on this application to inform the iterative process.
Multimodal			x	The application itself is presentable via the web and contains various multimedia elements including text and images. This case study was not published and had limited exposure outside of the evaluation phase of the project.
Open and accessible	x			Built in as a web application this Deep Map is accessible to all audiences. The code for this application is available as open source code on GitHub.
Spatial Humanities Criteria				
Relationship between space and time	x			Time is a key feature in the navigation of data in this prototype. The timeline function in the map allows for the exploration of spatial data in a temporal context.
Place built through multi-layered data	x			Each location in this Deep Map provides imagery and textual accounts through time to convey meaning to the audience.
Deep Mapping Criteria				
Spatial Exploration as a primary form of interaction	x			This application is explorable via both an interactive map and through the navigation of 3D panoramas. Searches are plotted onto the map and allow for exploration of the data across space.

Temporal context as a primary form of discovery	x			The timeline feature of this map allows for time to be a significant variable in the exploration of data returned by searches in this application.
Data from verifiable source for legitimacy	x			The primary data for this application is from the CLDW which is a verified historic literary source.
Layers fill targeted area	x			The target area of this map is centred on the Lake District, but also allows for navigation across the UK. The amount of data filling any area is dependent on individual searches, but the quantity of data within the CLDW provides a significant number of potential returns.
Accessible to broad audience	x			As a web application this prototype was accessible to a wide audience. The application itself has a standard interface design and its functionality is accessible to an audience with standard web browsing experience, but the focus on concordance matches and the advanced search features make this application more complex for uninitiated users.
Research Question Criteria				
Inform Research Process	x			This application provides the potential for new research processes to be discovered by the linking of various data sources in a spatial and temporal context. The 'Beautiful vs Sublime' map is an example of relationships that can be discovered to inform research.
Accessible and understandable				This prototype was primarily accessed during the evaluation phase of the project, but was also

				presented at some events with researchers and site authorities.
Researchers and Scholars	x			
Site Authorities	x			
General Public			x	While the public could understand the functional elements of this application, not all were familiar enough with the historical and literary content to fully understand the application's purpose. Additionally, the corpus linguistic concepts like concordance make this application more complex from the general public to understand.
Engaging				Feedback from users at presentations and demos showed this application was sufficiently able to engage audiences.
Researchers and Scholars	x			
Site Authorities	x			
General Public	x			
Repeatable	x			This Deep Map is easily repeatable with a text-based corpus that is marked up in the same fashion. This was demonstrated by the Holocaust Testimonies Processing case study (chapter 7), which was processed results have compatible tagging and would only require being added to this applications file system.
Expandable	x			This application is not limited in the scope of the data. New data can be added as long as it is correctly tagged to allow the geolocation of the searches. The only potential limitation to the addition of data is memory, if a dataset is too large it could be difficult for the system to

				effectively conduct searches through the data.
Total	19	0	2	

The prototype developed for this study was heavily targeted at academic audiences that could use this application as a tool for informing research. Due to this focus, the complexity of the search functions made the applications more difficult for the general public to engage with. The general public could still use the map, perform searches and still have a meaningful experience, but some features were not relevant to their exploration of the map.

4.7 Conclusion

The Analytic Deep Map provides an effective tool for users to engage with the CLDW. This tool provides a quick and easy method for researchers to prototype research questions that focus on the spatial distribution of locations or terms in the corpus. This quick prototyping allows for researchers to experiment with different queries and use meaningful results to start using their normal research processes, like the future pursuit of relationships of locations and terms in GIS.

This prototype is limited by the size of the corpus and the memory limitations of web applications. Searches of locations not represented in the corpus provide no results for users. While the addition of texts to this application is a simple process, the memory limits of the web application would prevent the use of significantly larger corpora. Beyond the practical limitations of this application, this prototype only allows for investigations with a significant focus on spatial features. This prototype does not implement corpus linguistic methods other than concordance and limited word frequency, and does not implement the vast array of spatial analysis tools available in GIS platforms. This application is a tool that can be used to prototype research questions in order to more quickly pursue avenues of interest using other technologies.

5 Case Study: Reading in Place Deep Map

5.1 Problem

The second solution to address the creation of a Deep Mapping application that incorporated the entire CLDW, was the Reading in Place Deep Map or Close Reading Deep Map, focusing on the user's interaction with the text. The NLP Deep Map described in chapter 4 was targeted at querying the CLDW and returning partial fragments of the texts that were relevant to the specific search. The NLP map is certainly a tool for distance reading, aimed at the analysis and visualization of queries into the texts and not on the full texts themselves. The Close Reading Deep Map is essentially the opposite of the NLP map, allowing all the texts in the corpus to be explored, but focusing on one text at a time.

The case study problem:

CSP1 – How can traditional linguistics methods be a primary method of interaction in a Deep Map?

CSP2 - How can a Deep Map be designed to be a tool for informing the research process using the full CLDW?

5.2 Objective of Solution

The goal of this case study was to create a Deep Mapping application that would allow for a closer look at each text in the CLDW. Close reading is a common practice for engaging with a particular text in literary studies (Alves 2015). This Deep Maps objective is to assist in the process of close reading and provide additional layers of data that might enhance the engagement of the text, and the understanding of places described in the text. The goal of this application was also to target a more casual audience than that of the NLP map, by providing a design which would be easily understandable to non-scholarly users. This application provides an interactive platform for the close reading of a text, like might be done by members of the general public interested in Lake District literature, or grade school students reading Lake District literature as part of their academic studies. The interactive nature of this prototype can enhance these users' experience of the text as they are reading and provides the potential for users to better understand and better connect with the literature.

The case study objectives:

CSO1 - Use the full CLDW

CSO2 - Use Linguistics concepts for data interaction
CSO3 - Provide layered data for each location to create a sense of place

5.3 Design and Development

The Close Reading Deep Map is a single page application, with the main page holding all the contents of the application. The main page is broken into three pieces (Figure 18). The top half of the page is a 360 Panoramic View using Google Street View. The bottom left of the page is a Notebook View, showing an image of a notebook, which becomes populated by a selected text from the CLDW. The bottom right is a Map View, using the API from the National Library of Scotland[34] to display historic Ordinance Survey maps of the United Kingdom.

When a user enters the Close Reading Deep Map, a random text is selected from the CLDW. The user can also choose to select the text themselves using a dropdown in the top of the Notebook View. Once a text is selected it populates the notebook with text. In order to create an immersive experience for the users, a background algorithm processes the raw text and splits it into sections that each fit onto a page in the notebook visualization. While these sections will look like a page in the notebook to the user, they do not necessarily reflect actual page breaks in the original text, as the sections are designed to fit to the notebook page image and adjust for screen size and resolution. Arrows on the sides of the notebook allow the user to flip through the text as if reading a book. Each location with an underlying enamex tag in the visible pages is presented in blue writing, setting them apart from the rest of the black text. The first location visible on the pages is by default considered the active location and is presented in an orange color. The user can use a Next and a Previous button at the top of the Notebook View to cycle through the visible locations in the text. Whichever location is currently active automatically changes the center point of the Map View to that location and changes the Panorama View to show the closest panorama to that location available from Google Street View.

[34] https://www.nls.uk/

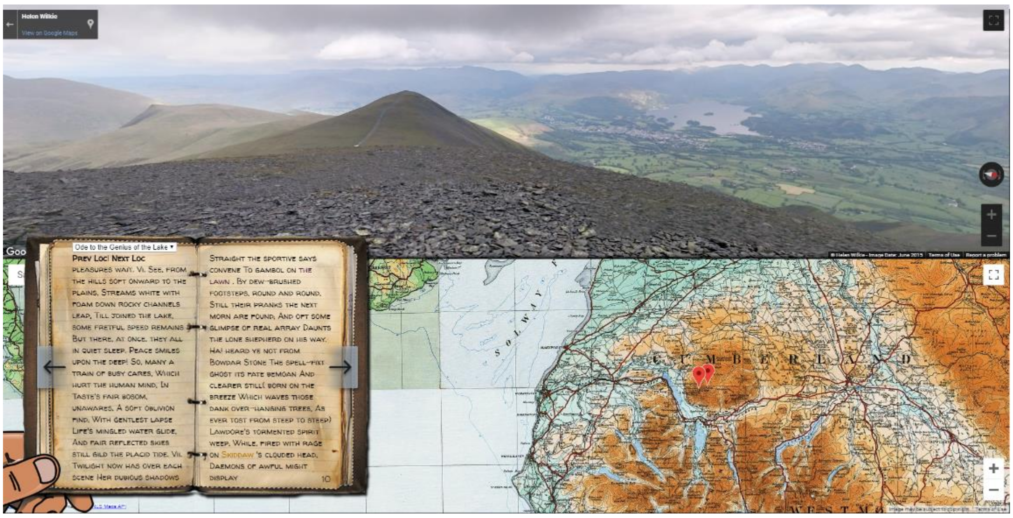

FIGURE 18 MAIN VIEW OF READING IN PLACE DEEP MAP

5.4 Demonstration and Audience

This prototype was aimed at casual readers in the general public and literary academics using close reading practices. This application was demonstrated at university seminars and at Lancaster University Community Days, gaining visibility with academics and the general public. While this application was well received by these audiences it did not get a lot of prolonged attention as it is aimed at reading a full literary text, and the environments of these demonstrations typically meant users would only engage with projects for a few minutes before moving on.

5.5 Evaluation

This section looks at the three categories of evaluation as defined in section 1.2. Each category has its own set of criteria and helps to provide insight into the effectiveness of this Deep Mapping implementation.

5.5.1 Design Evaluation

This section looks at how the problem and objective stated in 5.1 and 5.2 have been addressed by the design and development of the Reading in Place Deep Map.

Problem

CSP1 – How can traditional linguistics methods be a primary method of interaction in a Deep Map?

This application is targeted at enhancing the practice of close reading in the field of linguistics. The simple design and single page implementation of this prototype provide users with a functional environment in which to carefully read a text. The added panoramic imagery and map enhances the user's connection to places as they read about them in the text.

CSP2 - How can a Deep Map be designed to be a tool for informing the research process using the full CLDW?

In traditional close reading practices, users engage with a text independent of other data sources. By creating a link between the text and spatial data, relevant to the section of text the user is actively reading, it is possible for research derived from close reading to be enhanced and have greater consideration for the spatial context of the text.

Objectives

CSO1 - Use the full CLDW
This prototype gives users access to all the texts included in the CLDW. Users access each one individually, with no connections between the different texts, because the intent of this application is for users to focus on the careful reading of a single text.

CSO2 - Use Linguistics concepts for data interaction
The sole focus of this application is to enhance the close reading practice used in linguistic research. This was done by adding layered data in conjunction with access to the full text selected by a user. Any text from the CLDW can be selected by users and read in its entirety using this application, and as users read through a text they have access to additional data, through imagery and maps, to better contextualize the locations they read about.

CSO3 - Provide layered data for each location to create a sense of place
This application, like the previous prototypes brings different sets of data together to provide a multilayered engagement of place for users. The map, imagery, and text available for locations selected in the text come together to create a sense of place and provide the use the potential to develop a meaningful connection with the place.

5.5.2 Research Evaluation

This section of the evaluation looks at how this prototype helps to address overall research questions of this thesis.

RQ1 - How can Deep Mapping, in combination with spatial analysis, natural language processing, and corpus linguistic techniques, be used to inform the research process?

This Deep Mapping prototype was created as a tool to assist researchers, and other interested audiences, in the close reading of texts within the CLDW, while providing an enhanced experience focusing on visualizing locations in the text. Creating a link between the text and spatial data makes it possible for research derived from close reading to be enhanced and have greater consideration for the spatial context of the text.

RQ2 - What elements are required to best present historic data to various user groups in a way that makes it accessible and understandable? What visualization techniques best engage these different user groups?

The visual design of the prototype developed for this case study is intended to make users feel like they are holding a book and a map while standing at the locations they read about. While there are significant limitations to how much a web application can make users feel like they are physically at a location, the visuals of this map allow users to engage with the text and locations in a more active hands on situation. This visualization makes the content engaging for all audiences. The simplistic single page design of this application makes it easily accessible and understandable by all of the user groups targeted in this thesis.

5.5.3 Data Evaluation

This section looks at how well the prototype Deep Map fits within the definitions of Deep Mapping and related values and methods from the Spatial and Digital Humanities. An evaluation rubric was created for this thesis so that these elements could be assessed and compared. The rubric also includes a breakdown of this thesis's research questions in order to provide a base for comparison. The answers to each criterion in the rubric were derived from the evaluation survey, in person demonstrations, and seminar and conference presentations.

The categories and criteria in the evaluation rubric are discussed in chapter 1.2. Each criterion can be answered with a 'Yes', 'No' or 'Maybe'. 'Maybe' is used when the prototype partially addresses the criteria but does not fully fulfil the expected outcome of that criteria. For each criterion there is a comment field summarizing how this prototype addressed that criteria.

TABLE 12 READING IN PLACE EVALUATION RUBRIC

	Yes	**No**	**Maybe**	**Comment**
Digital Humanities Criteria				
Iterative Process	x			This Deep Map was developed through several phases and was improved upon using feedback from researchers involved in the project.

Experimental	x			As part of a PhD research project exploring the potential of Deep Mapping applications, this Deep Map is considerably experimental. Additionally with the lack of practice implementations of Deep Mapping applications, new applications like this one are experimental.
Collaborative	x			A multidisciplinary team of researchers provided feedback on this application to inform the iterative process.
Multimodal			x	The application itself is presentable via the web and contains various multimedia elements including text and images. This case study was not published and had limited exposure outside of the evaluation phase of the project.
Open and accessible	x			Built in as a web application this Deep Map is accessible to all audiences. The code for this application is available as open source code on GitHub.
Spatial Humanities Criteria				
Relationship between space and time		x		While the texts and the map are historic, this map does not emphasise the temporal components of the data.
Place built through multi-layered data	x			Each location in this Deep Map provides imagery and textual accounts to convey meaning to the audience.
Deep Mapping Criteria				
Spatial Exploration as a primary form of interaction	x			This application is explorable through the navigation of 3D panoramas and the selection of locations in the text. While there is no emphasis on spatial navigation

				though the map, this applications interaction and navigation is inherently spatial.
Temporal context as a primary form of discovery		x		This application does not use temporal navigation of the data as a means of discovery.
Data from verifiable source for legitimacy	x			The primary data for this application is from the CLDW which is a verified historic literary source.
Layers fill targeted area	x			The target area of this map changes via each location selected in the text. All the data displayed is related to the location that is currently in focus.
Accessible to broad audience	x			As a web application this prototype was accessible to a wide audience. The application itself has a standard interface design and its functionality is accessible to an audience with standard web browsing experience. This application was intended to be accessible to all audiences, including a younger demographic.
Research Question Criteria				
Inform Research Process	x			This application focuses on enhancing the practice of Close reading. Visualizing each location as it is explored in the text provides the potential for discovering new links between the text and spatially associated data.
Accessible and understandable				This prototype was primarily accessed during the evaluation phase of the project, but was also presented at some events with researchers and site authorities.
Researchers and Scholars	x			
Site Authorities	x			

General Public	x			
Engaging				Feedback from users at presentations and demos showed this application was sufficiently able to engage audiences.
Researchers and Scholars	x			
Site Authorities	x			
General Public	x			
Repeatable	x			This Deep Map is easily repeatable with a text-based corpus that is marked up in the same fashion. This application is primarily designed for literary texts that discuss real world locations.
Expandable	x			This application is not limited in the scope of the data. New data can be added as long as it is correctly tagged to allow the geolocation of the locations in the text. The only potential limitation to the addition of data is memory, if a dataset is too large it could be difficult for the system to effectively load large textual sources.
Total	18	2	1	

While this prototype presents historic literary texts and a historic map to users, there is no other functionality or navigational features that emphasize time. Given the focus on close reading and the attempt to make users feel like they are standing at locations they are reading about, this application did not necessitate using temporal features for the organization or navigation of data. Like the Crosthwaite Deep Map, this lack of temporal features does not exclude this prototype from being a Deep Map, as only a few iterations of Deep mapping definitions mention temporal relationships.

5.6 Conclusion

This case study aimed to create a prototype Deep Map that would enable users to read texts from the CLDW using close, or careful, reading practices common to literary research. The prototype allowed users to select a text from the CLDW and read the text in its entirety. Additional data is provided alongside the text enabling users to explore locations mentioned in the text as they are reading about them. This functionality allows users to see each location they read about in the text, explore the panoramic imagery, and understand the spatial context of the location via the map. As described in chapter 2.4.2, this Deep Map fits within the 'personal' category of Ridge et al.'s conceptual model and given the concentration on close reading of an individual text, is closely related to the idea of a spatial narrative. This prototype provides an engaging interactive experience for the reading of texts that was targeted at the general public (UG3), who might want to enjoy reading a text from the corpus, but also has value to literary scholars performing close reading practices in their research, as this application help put the writings in a spatial context.

6 Case Study: Tarn Hows – Automated 3D Reconstruction

6.1 Problem

This study looks at the creation of a 3D Deep Map, focusing on the National Trust site of Tarn Hows. As part of Lancaster University's *Geospatial Innovation in the Digital Humanities: A Deep Map of the English Lake District* project, we wanted to explore the concept of a 3D Deep Map and how it could be used to aid in the conservation of a historic site and provide a platform for educating visitors and the general public about the changing histories and landscape of the site. We chose to focus on Tarn Hows because it is a major tourist destination in the Lake District National Park, and the site has undergone significant changes to its landscape. For this case study we developed a series of scripts, for the Unity 3D gaming engine[35], that can take 2D elevation and vegetation maps, and automatically create a 3D visualization. These scripts were used to create the Tarn Hows Deep Map, with an interface for the application that allows for the navigation and comparison between historic periods.

This project investigates the use of Deep Mapping as the primary framework for creating an application to represent the changing historic landscapes of Tarn Hows. This Deep Map is a spatial representation of "the multiple histories of place, the cross-sectional stories of natural and human history (Maher 2014)." With the focus on place, Deep Maps can be seen as a multi-temporal chorography. Deep Maps are also intended to be 'open-ended exploration of a particular place' and have an 'emphasis on one particular user interaction: exploration.' An application that allows for this open-ended exploration, creates the potential for an environment where users can relate to the provided meaning attached to place, as well as derive their own from the experience (Ridge 2013).

The case study problem:

CSP1 - How can a Deep Map be made using a 3D environment?

CSP2 - How effective can a deep map be in providing insights for site authorities?

[35] https://unity.com/

6.2 Background

This project focuses on the National Trust site of Tarn Hows, a mountain lake located in the Lake District National Park. Tarn Hows is one of the most popular sites in the Lake District due to its picturesque scenery and easily accessible walking paths. The site today, is a single lake with an island in the middle, and abundant tree coverage across the site. While the site is known for its beautiful landscape, the site has undergone significant changes over the past 200 years.

In the early 1800s the site of Tarn Hows was comprised of three smaller natural tarns. The site was surrounded by farms and quarries and was used as a common grazing area. In 1862 the site became the property of the Marshall family, who had owned the local farms and quarries. James Garth Marshall endeavored to create a designed landscaped and dammed the stream flowing through the area, which created the single Tarn visible today. Marshall began a planting scheme as well, expanding various tree plantations around the site. Marshall's landscaping was designed to highlight features within the environment to create a more dramatic landscape[36].

Unfortunately, Marshall died before completing his vision of the site, but by the early 1900s the site was already a popular tourist attraction, mentioned in the writings of travel journals. In 1929 the Marshall family sold the land to Beatrix Potter, a popular Author and Artist, who later sold part of the land to the National Trust and left the rest to the Trust in her will[37].

The landscape continued to change over the years, with the tree plantations spreading out into once grassy areas, covering more of the lake's perimeter. Under the ownership of the National trust, footpaths and road access were added. Efforts were taken to reduce the impact of visitors on the site, including moving car parking lots to less visible areas. The site remained a major tourist destination in the area, reportedly drawing half a million visitors each year in the 1970s.

While the site is well known, and well visited, the changing historic landscapes of the site are difficult to show visitors without adding obstructive signs and plaques that would detract from the natural landscape. The National Trust must consider the modern aesthetic of the site as well as the historic landscapes when planning site maintenance.

[36] http://www.visitcumbria.com/amb/tarn-hows/

[37] https://www.nationaltrust.org.uk/tarn-hows-and-coniston/features/

6.3 Objective of Solution

The goal of this case study was to create a Deep Map that deviated from the map and marker style that the previous prototypes used. This case study was aimed at making a Deep Map using 3D visualization to experiment with Deep Mapping concepts outside of the standard 2D interpretations. Given that manual creation of 3D environments is time consuming and non repeatable by other researchers, one of the primary constraints of this case study was to make an automated process that would enable future users to implement the same methods.

The case study objectives:

CSO1 - Built in a 3D environment

CSO2 - Focus on automation, non-manual solution to building environment

CSO3 - Incorporate multimedia

CSO4 - Visualize the key historic periods of Tarn Hows

CSO5 - Allow comparison of time periods

6.4 Design and Development

This project uses the Unity 3D gaming engine, which allows for building an easily navigable 3D world. One of the key components to the 3D world in Unity is a Terrain. A Terrain is a plane object in the 3D space that is used as the ground. This plane is defined by a grid, with no data it is a flat object, and the changing of various values for each square in the grid creates the virtual landscape. Each grid square can be assigned values for attributes like height and color. Changing the colors of the Terrain is referred to as painting. Objects, like trees and grass, can also be assigned to each square. Some of the assigned objects can be 3D, like trees in the foreground, while others can be Billboards, which is a flat plane with a cut out image of the assigned feature. A billboard is used for elements like grass, or trees in the distance, and are intended to reduce the processing power required by the gaming engine for elements that don't need detail of a fully 3D object.

The scripts designed for this project use maps as inputs to perform their specific operation. These maps are Raster images, which is essentially a grid, where each square in the grid is a single pixel in the image and each pixel has a value that determines its color. The use of Raster images makes it relatively straight forward to bring the color data from the maps into Unity and assign those values to attributes in the Terrain grid.

FIGURE 19 TARN HOWS: DIAGRAM OF 3D ENVIRONMENT CREATION

Scripts

A series of scripts[38] were designed for the gaming engine Unity 3D, which could automatically build a 3D environment. Three scripts were made, the first was designed to create the topography of a terrain based on a 2D elevation map. The elevation script takes a traditional grayscale elevation map and uses the color value for each pixel on the map as the height value for the corresponding grid square on the terrain.

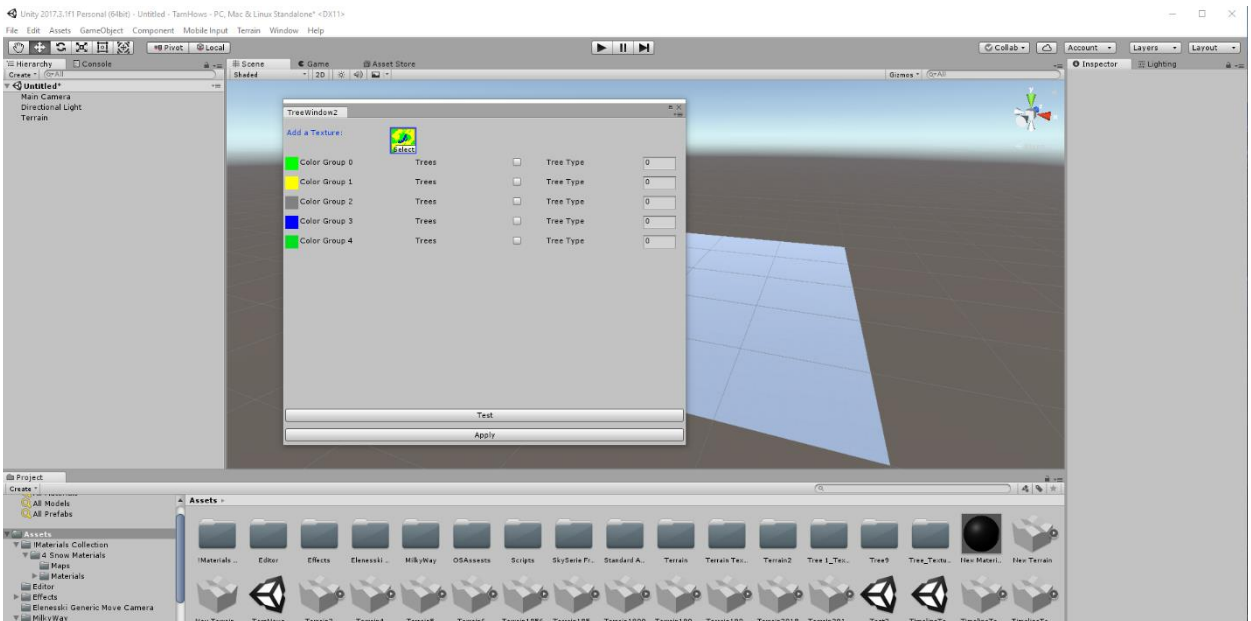

FIGURE 20 TARN HOWS: UNITY 3D INTERFACE AND VEGETATION SCRIPT TOOL

38 https://github.com/areinhold17/DeepmapScripts

The second script focuses on generating the minor features of the landscape. This script takes a 2D vegetation map, which is used to paint the terrain and create small details like grass. When a vegetation map is loaded into the interface for this script, the user is able to assign each color of the image to a different environmental detail they have added to the game engine. For instance, a user could assign the green pixels of the map to be grasslands, which could add a grass texture to the corresponding area on the terrain and add grass billboard objects to the area as well.

The third script is designed to generate tree cover. Like the second script, this one uses the same vegetation map as its input. The interface for this script allows the user to assign whether each color on the map should have trees, and assign which types of trees should be assigned to each color that trees are allowed. The script randomly places trees in assigned areas of the terrain, using an algorithm that spaces the trees based on a user specified distance range, ensuring that trees don't get too clumped together, or too far apart. This placement algorithm allows for a more natural looking tree cover and allows for control on the density of trees, to more accurately represent the real world landscape.

Tarn Hows Deep Map

To visualize the multiple historic landscapes of Tarn Hows, we compiled elevation and vegetation maps for three time periods of the site, 1856,1900, and 2018. We started with 2018, building an elevation map from LIDAR data available from the Government Digital Service[39], and augmented by satellite elevation data from the Government Digital Service for parts not covered by LIDAR survey. LIDAR data was used because it has a high resolution, about one meter, as compared to standard satellite data. The vegetation map for 2018 was based on satellite imagery.

The 2018 elevation map was used for the 1900 time period, as the site was a single Tarn, as it is today. The 1900 vegetation map was made using data from the Ordinance Survey (OS) map of the time period. While not as detailed as a satellite image, the OS map depicts where the various tree plantations were as well as the profile of the Tarn.

The 1856 vegetation map was also based on OS data, the 1856 OS map being the first to cover the area of Tarn Hows. This time period was before the merger of the three Tarns, and the OS map depicts the original Tarns. The 2018 elevation map was modified for this period using the limited elevation data and Tarn profiles available in the OS map.

With the elevation and vegetation maps prepared, we used the Unity scripts to automatically generate an environment for each period. We then built an interface

[39] https://gds.blog.gov.uk/

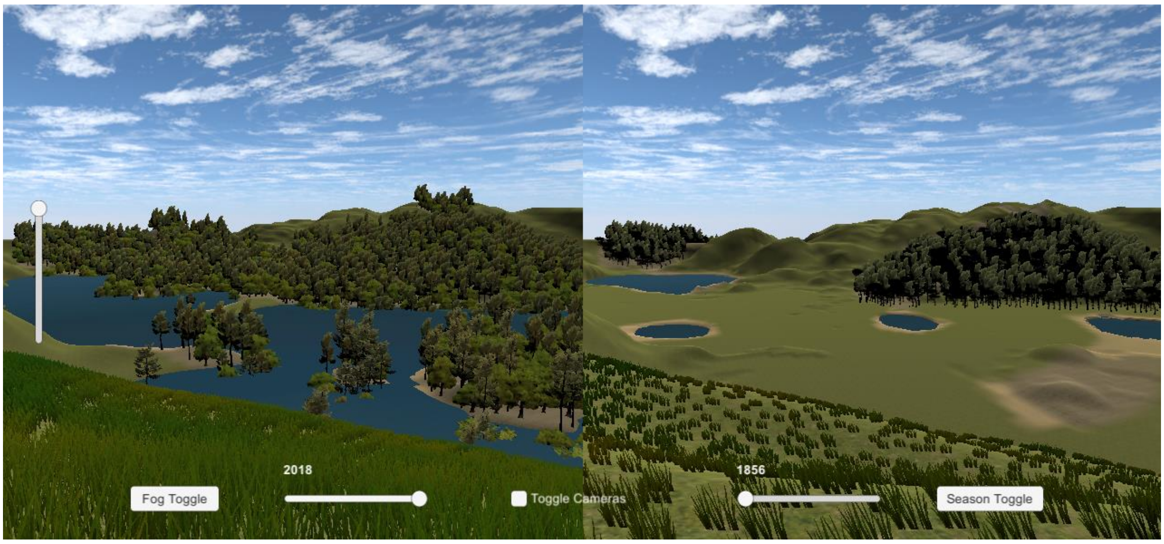

FIGURE 21 TARN HOWS: TIME PERIOD COMPARISON

that allows users to switch between time periods. The interface gives users the option of using one camera, to explore the site freely using traditional gaming interaction, or two cameras, to view two time periods side by side. When using the two camera view (figure 20), the cameras are synced, so any movement by the user is reflected in both views, allowing users to explore the landscape in multiple periods at the same time. A separate timeline slider for each camera permits the user to change and compare time periods at will.

The interface also includes the ability to change seasons in the landscape, which is done by switching out assets, the trees, grasses and paint textures, to reflect the desired season. Another slider in the interface gives users the option of changing the sky and the lighting level in the environment, and a toggle button allows for the addition of fog effects. These features provide enhanced realism to the virtual environment that could allow for more meaningful engagement with the application.

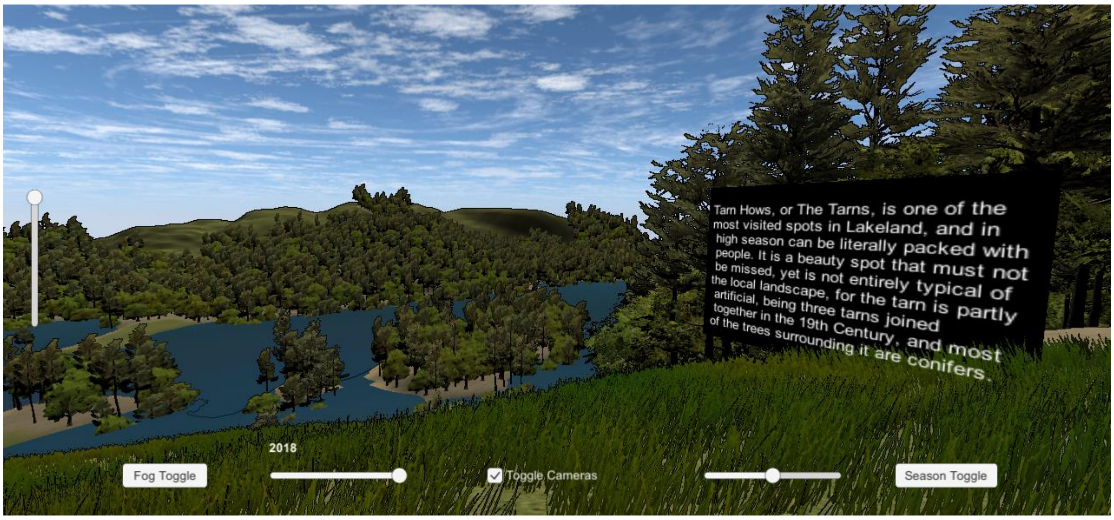

FIGURE 22 TARN HOWS: VIEW WITH MULTIMEDIA EMBEDDED

As this application is built in the Unity game engine, any number of multimedia elements can be added to the environment in order to enhance the user's engagement (figure 21). As an example, we added a sign which explains a little about the site, that users can read as they overlook the lake. This type of media placement is important because it provides a means of disseminating interpretive data that cannot be placed in the real world environment without detracting from the natural beauty of the site.

6.5 Demonstrations for formative feedback

This work was presented to the National Trust and the Lake District National Park Authority (LDNPA) in two separate meetings focusing on how Deep Mapping applications can be used in tourism and conservation. The National Trust was interested in the Deep Map as Tarn Hows is a National trust site. The LDNPA had some interest in the Deep Map, but the projects run by the individuals present were typically aimed at a regional level and the Deep Map focuses on site level representation. This work was also presented at the Geospatial Seminar series, and got good feedback from the attendees, especially from an archaeologist with Oxford archaeology, interested in the potential of this work for the visualization of ancient landscapes based on archaeological records.

This work was presented at the Visual Heritage 2018 Conference in Vienna (Reinhold 2018) and at the Spatial Humanities 2018 Conference in Lancaster.

6.6 Evaluation

The Tarn Hows Deep Map application that we created provides a tool for both education and public interaction, as well as a means of site conservation and management. This application can be used as a platform for educating the public about the layer history of the site. With embedded multimedia and the functionality to move between time periods, the Deep Map provides a means of 'layered storytelling' and allows for the open-ended exploration of the place.

For site authorities, this application can help visualize past environments to inform site maintenance, and visualize the future landscape based on planned changes. With sites that have had a rich history of landscape change it can be difficult for authorities to plan maintenance of the site. In the case of Tarn Hows, the modern environment is an iconic tourist destination, but today's landscape differs significantly from the past. Authorities must determine what trees can be cut back to better reflect the historic plantations of the Marshall family, while not removing too much that might negatively affect the picturesque views and visitors' engagement with the natural landscape. The Deep Map also provides a means to visualize future plans for the site, allowing authorities to gain feedback on proposed changes without having to modify the real world landscape first.

The automatic generation of a landscape, by the scripts written for this project, provides the means to quickly create visualizations of modern and historic environments. While there is an issue of accuracy, such as individual trees do not necessarily reflect the positions of their real world counterparts, this method allows for the rapid creation of an environment, even with limited data.

6.6.1 Design Evaluation

The case study problem:

CSP1 - How Can a Deep Map be made using a 3D environment?

The Tarn Hows Prototype shows that a Deep Map based on a 3D environment qualifies under the definition of Deep Mapping, given that it is spatially explorable, allows for embedding of media, and has a multilayered design. The gaming engine underneath this application offers the functionality for creating a freely explorable world, a key element in Deep Mapping. With the various time periods represented by the 3D environments and other embedded multimedia, the application provides a layered approach to presenting the data.

CSP2 - How effective can a deep map be in providing insights for site authorities?

This prototype provides a platform for assisting site authorities with the conservation of natural sites. The ability to visualize past landscapes as well as modern or planned landscapes can give authorities insight into how vistas change. Site authorities for Tarn Hows talked to as part of demonstrations of this prototype noted that there is a continuous concern about the management of tree removals and site maintenance. With the past landscaping phases the site has undergone, conservators have to determine what phase the modern landscape should look like. This application is a tool that conservators can add maps of planned changes in order to compare them to the modern and past environments. This comparison can highlight changes to iconic vistas that they might not wish to change.

Objectives

CSO1 - Built in a 3D environment

The Unity gaming engine provided an easy environment to build this application. The scripting part of the Unity engine allowed for 2D images to be converted to 3D surfaces and to populate those surfaces with 3D assets. That scripting enabled the automated building of environments based on altered historic maps.

CSO2 - Focus on automation, non-manual solution to building environment

Automation was a key element to the design of this application. The manual creation of a 3D environment can be very time intensive and required significant amounts of professional expertise. If this application was to be a tool for other researchers to use in the future, a manual process would make it inaccessible to a large portion of interested researchers. The scripting function of Unity provided the optimal solution to this issue because scripts could be created to perform the required actions and then those scripts can be shared with others to enable them to build similar Deep Maps. Implementing premade scripts is relatively easy and can be done by users without expertise in 3D modeling with basic instructions. The 2D maps that are required to run the scripts only require some image editing software to create and are likely to be the result of other planning or research processes.

CSO3 - Incorporate multimedia

The free structure of the 3D environment allows for the easy addition of additional assets. Text and Imagery can be added as basic object within the world, and other multimedia like videos can be added with a little more work. With the heavy emphasis on exploration in this type of application, the positioning of multimedia assets around the world allows for users to discover content as they openly wander the world.

CSO4 - Visualize the key historic periods of Tarn Hows

CSO5 - Allow comparison of time periods

CSO4 and CSO5 overlap and are addressed together here. In addition to the focus on automation of the creation process, this application's key interactive feature is the experience of various time periods. Each time period represented in the application can be compared to one another or explored individually. The scripts created for this case study allow for the quick creation of environments for various time periods, only requiring vegetation and elevation maps that represent the time periods of interest. The timeline interface in the prototype enables users to compare the automatically generated 3D environments of the different time periods side by side. In the comparison view users can move through the environment and their movements are reflected in both time periods being viewed, providing users the ability to actively engage with and explore changes to the landscape over time.

6.6.2 Research Evaluation

This section of the evaluation looks at how this prototype helps to address overall research questions of this thesis.

RQ1 - How can Deep Mapping, in combination with spatial analysis, natural language processing, and corpus linguistic techniques, be used to inform the research process?

This Deep Map provides a tool for a few different types of analytic and research processes. As discussed above, this application is a viable tool for conservation and related research into the management of historic landscapes. This application also provides a tool for generating 3D data that can be used for spatial analysis tasks commonly performed in digital humanities GIS projects, like viewshed analysis.

RQ2 - What elements are required to best present historic data to various user groups in a way that makes it accessible and understandable? What visualization techniques best engage these different user groups?

This application is viable among all the user groups of interest to this PhD's research. The application was well received by all audiences. The visualizations present the historic periods in a meaningful way that engages audiences in the material. The 3D format is widely accessible and understandable to all audiences, with young audiences having an easier time engaging with the application.

The 3D format of this application was among the most engaging features used to visualize historic content out of all the prototype maps designed for this PhD. 2D maps are often difficult for some users to fully understand and appreciate because they present data in a manner that can be too abstract for some. The 3D visualization is familiar to all audiences because it can be explored as an interactive environment or it can be presented as 2D images.

6.6.3 Data Evaluation

This section looks at how well the prototype Deep Map fits within the definitions of Deep Mapping and related values and methods from the Spatial and Digital Humanities. An evaluation rubric was created for this thesis so that these elements could be assessed and compared. The rubric also includes a breakdown of this thesis's research questions in order to provide a base for comparison. The answers to each criterion in the rubric were derived from the evaluation survey, in person demonstrations, and seminar and conference presentations.

The categories and criteria in the evaluation rubric are discussed in chapter 1.2. Each criterion can be answered with a 'Yes', 'No' or 'Maybe'. 'Maybe' is used when the prototype partially addresses the criteria but does not fully fulfil the expected outcome of that criteria. For each criterion there is a comment field summarizing how this prototype addressed that criteria.

TABLE 13 TARN HOWS EVALUATION RUBRIC

	Yes	No	Maybe	Comment
Digital Humanities Criteria				
Iterative Process	x			This Deep Map was developed through several phases and was improved upon using feedback from researchers involved in the project.
Experimental	x			As part of a PhD research project exploring the potential of Deep Mapping applications, this Deep Map is considerably experimental. Additionally with the lack of practice implementations of Deep Mapping applications, new applications like this one are experimental.
Collaborative	x			A multidisciplinary team of researchers provided feedback on this application to inform the iterative process.
Multimodal	x			The application itself is presentable via the web and contains various multimedia elements including 3D visualization, text and images. This case study was published and presented upon at two conferences.
Open and accessible	x			This application was built using the Unity gaming engine which allows for publishing to the web, mobile devices, and standalone desktop environments.
Spatial Humanities Criteria				
Relationship between space and time	x			One of the key features to this application is the ability for users to explore visualizations of different time periods and also compare them side by side.
Place built through multi-layered data	x			This Deep Map contains several layers of temporal visualizations as

				well as embedded multimedia like text and images, that contribute to building a sense of place
Deep Mapping Criteria				
Spatial Exploration as a primary form of interaction	x			This application is a 3D gaming experience and allows free movement through the environment. Users can explore the open world of the application and discover embedded multimedia in their exploration.
Temporal context as a primary form of discovery	x			The temporal element of this application is a key function of discovery with the interface allowing users to explore and compare environments set in different time periods.
Data from verifiable source for legitimacy	x			The primary data for this application is from satellite data and Ordinance Survey maps.
Layers fill targeted area	x			The target area of this map is a 1km square area around Tarn Hows. All locations within this area have a visualization for each specified time period.
Accessible to broad audience	x			As a web application this prototype was accessible to a wide audience. The application's interface is as simple as possible to provide the functionality required. The controls are standard gaming controls for a 3D environment (WASD) and also allows for the arrow keys to be used for movement which provides both younger and older audiences an understandable method of interaction.
Research Question Criteria				
Inform Research Process	x			This application focuses on rendering historic environments

				and allows comparisons across time. This functionality can provide insight into research, education, and conservation.
Accessible and understandable				This prototype was primarily accessed during the evaluation phase of the project, but was also presented at some events with researchers and site authorities.
Researchers and Scholars	x			
Site Authorities	x			
General Public	x			
Engaging				Feedback from users at presentations and demos showed this application was sufficiently able to engage audiences. This application was best received by younger audiences in the demos as they enjoyed the exploration of 3D environments.
Researchers and Scholars	x			
Site Authorities	x			
General Public	x			
Repeatable	x			This Deep Map is easily repeatable with any site that has 2D maps showing vegetation cover over various time periods.
Expandable	x			This application is not limited in the scope of the data. New visualization layers can be added as long as there is a 2D map to use as a base. New multimedia can easily be embedded into the map as well.
Total	21			

The Tarn Hows Deep Map was the only prototype developed for this thesis that fulfilled all the criteria used for evaluation, as seen in Table 13. The focus of this application on visualization, exploration and temporal change fulfilled criteria that other prototypes did not. Through demonstration and evaluation through the survey, this map proved to be interesting and engaging to all users. While some users had minor difficulty with accessibility at first, finding it difficult to understand how movement worked within the 3D environment, it did not require a significant amount of direction to bring users up to speed on that function of the application.

This application, through the use of the scripts, is easily repeatable and expandable to new locations and new time periods. The limitation of this prototype is that it was designed to recreate natural environments, not man-made ones. There is no functionality within the scripts to create 3D structures, like fences, houses or dams. While these elements can be added to the automatically generated environments produced by the scripts, developers would be required to manually create those assets. This lack of functionality for the inclusion of man-made structures in the automated process is not a failure of this case study, it is merely outside of the scope of what this prototype and the corresponding scripts were designed to accomplish.

6.7 Conclusion

This case study has allowed for the creation of a tool for quickly visualizing the past. Normal methods of 3D modeling are time consuming and require specialized expertise in order to construct 3D environments. The scripts developed for this case study automate the process of generating a 3D environment, potentially requiring only minutes to go from basic 2D imagery to a 3D landscape. The tool provides a means of educating users about the multiple histories of a site, in addition to providing a tool that site authorities and other organizations can use to visualize modern and historic landscapes. This application shows that a 3D Deep Map delivers an explorable visualization and can represent the multiple histories of a place in an engaging way. In the future we will be working with site authorities to use this tool to aid in the conservation of Tarn Hows and other sites and create educational exhibits to help disseminate historic data to the public.

7 Case Study: Holocaust Testimonies Processing

7.1 Problem

The bulk of this thesis's research focuses on the Lake District, using the CLDW as a primary dataset. The CLDW, while varied in its composition, was collected to focus on texts that emphasize location and therefore would be an effective dataset for geoparsing and spatial analysis and visualization. Due to the targeted nature of the CLDW, the question was asked, can a dataset, not collected with an emphasis on location, be processed in order to produce an output that would make it a usable dataset for the Deep Mapping applications previously discussed? Researchers partnered with Lancaster University were interested in the techniques used by the Geospatial Innovations project to process unstructured texts, to process Holocaust Testimonies and interviews collected by the United States Holocaust Memorial Museum (USHMM)[40] and University of Southern California Shoah Foundation Visual History Archive (VHA)[41]. The testimonial data set is comprised of approximately 5676 files (525 USHMM, 5151 VHA), a much larger corpus in comparison to the CLDW at 80 files. These files are transcripts of audio and video interviews done by the two organizations with survivors of the Holocaust. The USHMM files contain one interview per file, while the VHA interview are split between around 7 files on average. The Holocaust testimonial corpus represents approximately 1360 individual interviews. This large quantity of text files requires an automated approach to processing in order to bring the data into other tools for meaningful analysis.

7.2 Objective of Solution

This case study looks at how the corpus linguistic techniques used by the Geospatial Innovations project could be used to process the corpus of Holocaust Testimonies. The goal was to create a processed output that would identify locations within the testimonies and mark them with enamex tags. The researchers working with the testimony data were also interested in marking up the texts to identify questions asked by the interviewers and answers given by the interviewees, identify the name of the

[40] https://www.ushmm.org/

[41] https://sfi.usc.edu/what-we-do/collections

interviewee and their gender, and markup metadata to make it usable by other applications.

7.3 Design and Development

The Holocaust testimonies corpus is made up of two datasets. The first dataset is from the United States Holocaust Memorial Museum (USHMM)[42]. Each interview in the USHMM dataset starts with a section of metadata related to the file and interview, followed by a series of questions and answers, using a Question/ Answer or Q/A label format:

> Question: Well, can you tell me the names of your parents and brother?
>
> A: Uh-huh. My parents name is Noah Blass, my father. Then my mother was Bluma Blass. Her maiden name was Goldvasser, and what -- what else?
>
> Q: Your brother's name?
>
> A: Oh, my brother's. My brother's name was Moshe Ben Blass. It's Hebrew names, and they were translated int-into Monyek Blass.
>
> Q: And was he younger or older than you?
>
> A: He was a year and a half younger than I am, and he died unfortunately, in a camp near Auschwitz, a camp for men, Brunow. He was there with my father, and after the war I found out that they both died there. That's about all I heard about them.

The second dataset in the Holocaust testimonies corpus is from the University of Southern California Shoah Foundation Visual History Archive (VHA)[43]. Each file in this dataset contains only the interview, with no associated metadata included in the file. In its raw form each file is mostly illegible, using a 'span' tag to mark every word and provide a value for each unique word:

```
<transcription><p><span m='35260'>INT 1:</span> <span m='35384'>With</span> <span
m='35758'>name,</span> <span m='36290'>date,</span> <span m='36750'>location.</span> </p><p><span
m='37620'>INT 2:</span> <span m='37837'>OK.</span> </p><p><span m='39800'>EJ:</span> <span
m='39910'>You</span> <span m='40020'>are</span> <span m='40200'>going</span> <span
m='40410'>to</span> <span m='40520'>say</span> <span m='40850'>that?</span> </p><p><span
m='41268'>INT 1:</span> <span m='41477'>You</span> <span m='41686'>can say that.</span>
```

These span tags were a byproduct of work done to link each word of the text with the corresponding word in the audio or video files. Since the aim of this processing was unrelated to the audio and video file this information was deemed unnecessary.

The interview is formatted with INT 1 or 2 labels for the interviewers and the Interviewees initials as labels to denote who is talking. The structure of the VHA interviews was more of a casual conversation rather than the question answer style of the USHMM testimonies. When the extra tags are stripped out of the file the text becomes more legible:

42 https://www.ushmm.org/

43 https://sfi.usc.edu/what-we-do/collections

> INT 1: I think it'll pick up better if Erika does it. We can all do it. It doesn't matter.
>
> EJ: But, name-- that my name is so and so and I live at this and this address . Is that what you want?
>
> INT 2: I wouldn't give the address. Just the city.
>
> EJ: OK, and the date. Today is the 11th?
>
> INT 2: Right.
>
> INT 1: Yes.
>
> EJ: OK. Yes, now?
>
> INT 2: Well, just a minute, because in the instructions.
>
> EJ: OK, fine. Then you start.
>
> INT 1: We are rolling and go.
>
> INT 2: OK, my name is Barbara Gordon. I'm the interviewer for Erika Jacoby in Valley Village, California. Today is July 11, 1994. Erika, tell us about your childhood.
>
> EJ: I was born in Miskolc, Hungary, which is Eastern Europe, 66 years ago. My mother described the day. It was very beautiful. It was May 1, and she delivered me without any help...

In order to process these files, both for geoparsing and to tag each question and answer, a custom application was designed. While the Edinburgh Geoparser was effective on the CLDW texts, the formatting of the Holocaust testimonies, specifically the raw VHA files, would not produce usable results, due to elements like the XML tags and the interview formatting ("INT1:","EJ:") that made the tokenization step in the NLP process return incorrect results. The Edinburgh Geoparser also would not include the additional markup required by the researcher, like labeling names, genders and what sections were questions and answers. The custom application created for this task is called the Interview Parser[44] and was written in Java. Java was chosen because several robust opensource NLP libraries are available in Java.

The Interview Parser is a processing application and has no graphical user interface. The parser begins by loading in the text for an interview and does an initial cleaning of the text, stripping out elements like the <span> tags in the VHA dataset or header and footer information that is present when loading pdf files into Java. The processor strips these pieces out of the text using Regular Expression (regex) statements, which provide the processor with specific rules:

Code:

```
        regex = "<span m='\\d*'>([^<]*)<\\/span>";
        replacement = "$1";
        text = replace(text, regex, replacement);
Meaning:
        Look for Every <span> </span> pair.
```

[44] https://github.com/areinhold17/Holocaust_Code

```
Remove the <span> tags and only keep the interior
contents.
```

The parser then looks for every instance of labels that denote each participant, in order to begin inserting xml tags to mark them. Given the variation of labels used in the VHA testimonies, the algorithm for identifying participants separates the labels into different groups.

TABLE 14 EXAMPLES OF LABELS AND ASSOCIATED XML TAGS

USHMM		**VHA**	
Label	Marked up with	Label	Marked up with
Question:	<question>	INT:	<question>
Q:	<question>	INT1:	<question>
Answer:	<answer>	INT2:	<question>
A:	<answer>	CREW:	<otherspeaker>
		UNIDENTIFIED SPEAKER:	<otherspeaker>
		(Interviewees Initials):	<answer>

With each instance of each participant label identified the text can be marked up with the relevant tags. A starting tag is added before each label, ex. "<question>Q:", and then an algorithm determines if a label follows another label then a closing tag should be added: if an "A:" appears as the next label after a "Q:" the algorithm will insert the closing tag as well as the starting tag "</question><answer>A:".

```
<question>INT: Where did you find her? </question><otherspeaker>CREW: Is she going to be interviewed, do you know? </otherspeaker><otherspeaker>CREW: You might want to go back over that picture because I was still running time. </otherspeaker><answer>SH: No, no. I didn't find her. Freddie Diamond came to me to visit, to Wiesbaden.  He said, Sieg, I met a girl by the name Halbreich. How does she look? </answer><question>INT: OK. </question><answer>SH: This is Oswiecim, later Auschwitz, before the war. </answer>
```

The next step in the parser is to run the text through NLP to identify people and places. The Interview Parser uses the Stanford CoreNLP[45] library to perform the main NLP and NER operations. This library takes the raw text as an input and outputs a list of all the Tokens in the text and provides their part of speech, the named entity tag. The parser then runs through all the tokens to find those with a relevant NER tag. Words with a NER tag of Person are tagged with a '<person>' tag and added to a list in the tagged metadata so that the researchers can see who is being talked about in the

[45] https://stanfordnlp.github.io/CoreNLP/

interview. The list of persons was then used to identify the full name of the interviewee if it was not stated at the beginning of the interview, which was common in the VHA dataset. The parser then used the interviewee's name and an external library[46] to identify the interviewee gender and add that data to the tagged metadata. Gender was split into four categories, Male, Female, Unisex and Unknown. Names that were typically a single gender, male or female, were assigned those labels, while names, like 'Alex' fell under the unisex category. Names that could not be found by the external library were assigned the unknown label, like the name 'Sulamif.'

Words identified as a location by the NER process are then run through the custom geoparser process. The geoparser compares the location against a database of gazetteer data. Three types of matches are used: full match, outer partial match, and inner partial match. A full match means the location listed in the gazetteer is an exact match to the searched location. An outer partial match means the searched location is contained within the gazetteer match. The inner partial match means the gazetteer match is contained within the search location (Table 15).

TABLE 15 EXAMPLES OF MATCH TYPES

NER Location	Gazetteer Location	Match Type
Washington	Washington	Full Match
Washington	Washington D.C.	Outer Partial Match
Washington D.C.	Washington	Inner Partial Match
Windermere	Windermere	Full Match
Windermere	Winder	Inner Partial Match
Winder	Windermere	Outer Partial Match

The geoparser finds every match for each search term and creates a confidence score for each based on the match type and the Levenshtein distance algorithm which quantifies the differences between two words (Gonzalo 2001). The Levenshtein distance algorithm calculates the number of changes required for input 1 to equal input 2. The words 'Lake' and 'Lakes' would receive a score of 1 because 1 character, 's', would need to be added to 'Lake' to make 'Lakes', and the words 'Lakes' and 'Lamps' would get a score of 2 because the two characters, the third and fourth, would need to be changed for the words to be the same , see Table 8.

TABLE 16 EXAMPLES OF THE LEVENSHTEIN DISTANCE SCORE

Input 1	Input 2	Levenshtein Distance Score	Changes
Washington	Washington	0	No Changes

[46] https://gender-api.com/en/

Washington	Washington D.C.	5	5 char added
Washington D.C.	Washington	5	5 char removed
Washington's	Washington	2	2 char removed
Windermere	Windermere	0	No Changes
Windermere	Winder	4	4 char removed
Winder	Windermere	4	4 char added
Winder	Winter	1	1 char changed
Winder	Water	5	4 char changed, 1 removed

A full match receives a confidence score of 100, while other matches use an inverse of the Levenshtein Distance score to calculate a confidence score, since the less changes required to match the two locations should indicate a more reliable match. After this process the locations identified by NER are then tagged in the text with the coordinate information from the gazetteer using the most confident match and all other matches for a location are logged into a database for reference or correction by the research team.

The geoparser process is the final segment of the Interview Parser and once complete returns a fully marked up text:

<interviewee>Siegfried Halbreich</interviewee><gender>male</gender>

<person>Clara Firestone</person><person>Keren Kayemet</person>[SIDE CONVERSATION] <otherspeaker>CREW: Rolling. </otherspeaker><question>INT: The name of our survivor today is <person>Siegfried Halbreich</person>-- spelled S-I-E-G-F-R-I-ED H-A-L-B-R-E-I-C-H. The interviewer's name is <person>Clara Firestone</person>. We're interviewing at <enamex wordUID="WRD1528766710595NNm3" long="-118.24368" lat="34.05223" type="location" gazref="GAZ15287622545880ycb" name="Los Angeles" conf="100.0">Los Angeles</enamex>, <enamex wordUID="WRD1528766713644TfjG" long="0.083333" lat="54.05" type="location" gazref="GAZ1521655590814GtTt" name="California" conf="100.0">California</enamex>. And the date today is July 12, 1994. Sig, I'd like to start with having you tell us a little bit about your life prior to the occupation-- prior to the beginning of the war-- where you were born, what your family life was like, your family members, et cetera. </question><answer>SH: I was born November 13, '09, in <enamex wordUID="WRD1528766716628MUYc" long="15.07872" lat="53.02128" type="location" gazref="GAZ1521655896410oFUT" name="Dziedzice" conf="100.0">Dziedzice</enamex>. It's a small town near <enamex wordUID="WRD1528766720012zsbx" long="19.04686" lat="49.82245" type="location" gazref="GAZ1521655896410oFUT" name="Bielsko-Biala" conf="100.0">Bielsko-Biala</enamex>. When I was born, this part of the country belonged to <enamex wordUID="WRD1528766723096uluF" long="-2.616944" lat="51.596389" type="location" gazref="GAZ1521655590814GtTt" name="Austria" conf="28.57142857142857">Austria</enamex>. When the war broke out, first of all, I was born-- I was seven months a baby. The doctor gave up. He said I would not live. So my grandmother bundled me up and took me to her hometown. And I was a raised for almost four years by my grandparents. I didn't know my parents. They came from time to time to visit us. But there were other sisters and brothers who came. So I did not. </answer><question>INT: They are readjusting the lens. </question>…

7.4 Demonstration

The results of the Interview Parser for the Holocaust dataset were provided to researchers at the University of Maine. The processed interviews are now compatible with corpus linguistic tools like wmatrix, and will allow researchers to query the corpus to identify themes and trends across the entire dataset. The processed interviews can also be used as a data source for a Deep Mapping application.

7.5 Evaluation

Unlike the other case studies in this thesis, this case study did not involve the creation of a Deep Map prototype. Since there is no prototype to evaluate the design and data portions of the case study evaluations does not particularly apply. The research section of the evaluation is still applicable because this case study, even without a prototype, makes some contribution to the resolving of the research questions.

7.5.1 Research Evaluation

This section of the evaluation looks at how this case study helps to address overall research questions of this thesis.

RQ1 - How can Deep Mapping, in combination with spatial analysis, natural language processing, and corpus linguistic techniques, be used to inform the research process?

This case study shows that NLP is a critical method for the inclusion of text based data into Deep Maps. Without NLP and geoparsing processing of text, computer applications would not be able to effectively map locations in the text without significant curation from developers. The geoparser used for this case study was published in 'A deeply annotated testbed for geographical text analysis' (Rayson et al 2017). That work compared three NLP libraries including the Stanford CoreNLP library used for this case study. The results of the statistical analysis of the Stanford library showed an average precision of 0.90 and a recall of 0.57. This high precision but low recall means that the system returns fewer results, missing some instances of locations, but the results returned are much more likely to be correct, so the system is unlikely to mistake person or organization names for locations. This means the results of this case study's processing are likely (90%) to produce correct locations, but it also means there is potentially more work to be done to identify locations missed by the processor.

7.6 Discussion

The interview parser provides a pipeline for marking up a corpus in a manner that makes it usable as a spatial dataset. While other NLP platforms and geoparsers are capable of processing unstructured text in a similar fashion, it was necessary to build a custom application in order to tag the specific elements of interest to the partnered researchers. A custom application also allowed for the easy use of user defined gazetteers to provide more accurate historic and temporal spatial coordinate data specific to the dataset.

Limitations and Future Work

The Interview parser was specifically designed to process the Holocaust testimony data, so it would require some modification to handle another dataset. The benefit of its design is that most of the components are modular. Making it compatible with a new dataset would be as simple as defining a new set of custom rules, like regex section that cleans the Holocaust dataset, which is optional and choosing which modular processes to run on the new dataset. The modular processes include people and location identification though NLP and NER, gender identification of names, and location matching to a gazetteer.

8 Conclusions

The final chapter of this thesis summarizes the research undertaken and its findings. This chapter reviews the research questions stated at the beginning of this thesis and considers how well they have been addressed by the research conducted. Then there will be a discussion of the contributions this research makes to the field and its broader impact. Finally this chapter will discuss the potential paths future research on this subject could take.

8.1 Summary of Work

The focus of this research was to implement and evaluate Deep Mapping applications based on data from the Corpus of Lake District Writing (CLDW). This was done through a series of case studies in order to create unique applications targeted at engaging the different user groups and addressing the research questions of this thesis as described in section 1.5. Four case studies were undertaken in which a prototype Deep Map was created to address an asserted problem for each case. The fifth case study, the Holocaust Testimonial processing (chapter 7), was undertaken in order to test the methods of data preparation, used for the first four case studies, on a data set not related to the Lake District.

The first case study, the Crosthwaite Deep Map (chapter 3), was aimed at creating a small scale Deep Map with a selection of texts from the CLDW. Given the flexible definition of Deep Mapping and the large collection of texts in the CLDW, a small and focused Deep Map case study was determined to be a good starting point of this research in order to explore and experiment with the concept of Deep Mapping. This case study focused on a historic map illustrated by Peter Crosthwaite of Derwentwater, a large lake in the Lake District. The map depicted picturesque locations described by the author Thomas West in the 1700s. The focus on this map and set of picturesque locations gave the Deep Map a bounded area to concentrate on and allowed the CLDW to be scaled down to only include texts that described the same set of locations. The prototype map developed for this case study was based on a standard web map interface, using markers to mark the locations of the picturesque stations. The map interface helped satisfy the Deep Mapping requirement of spatial exploration and allowed for the historic map to be overlayed into the interactive space. The markers on the map could be selected and would bring users to a new view focusing on the selected location (station view). This view was based on a 360 panoramic image to provide users with engaging imagery and a sense of being at the location. The view included excerpts from a subset of the CLDW that mentioned the selected location, as well as imagery of artistic renderings (drawings and paintings) done of the site. These elements were aimed at providing users with enough information that they could establish a meaningful connection with the location. This prototype also included a text view that was linked to both the map and the station view, allowing users to read through the original texts and freely explore the collection of data in whatever manner most engaged them via the three different views.

The first prototype met the requirements of Deep Mapping and proved to be an effective implementation, meeting the requirements of Deep Mapping as defined in the evaluation and providing a launching point for informing further research related to this thesis. This case study was presented at multiple public events, university seminars and at two conferences, COSIT 2017 and Digital Pasts 2018. The case study was the focus of a publication for the COSIT 2017 SPHINx workshop.

With the completion of the Crosthwaite Deep Map, the next step in the investigation of Deep Map was to expand the focus to include the complete CLDW. As this was an exploratory process, two different approaches to expanding the scope were tried, looking at the linguistic concepts of close and distant reading.

The second case study, the Analytic Deep Map, was aimed at using the entire CLDW to enable distant reading practices and incorporate corpus linguistic methods. This application used the corpus linguistic technique of concordance matching to perform searches of user defined terms in the CLDW. Users could search single terms, compare terms, or search by location, given them a variety are methods for querying the corpus. The prototype, like the Crosthwaite application is based on a web map interface. This interface allows search via the top navigation panel and displays results on the left panel in text form and as markers on the map. The markers that are places on the map are representative of locations found within the concordance match, so a search of the word 'waterfall' would show up on the map if a real world location was mentioned within 10 words (or based on the user defined concordance limit) of 'waterfall.' Selecting a marker on the map would allow users to access the location view, similar to the station view from the Crosthwaite map. The location view showed a panoramic image and displayed new results in the left panel that represented every instance of the selected locations mentioned within the CLDW. Like the Crosthwaite application, this prototype included a text view as well that allowed for multiple ways to engage with the corpus. While the text view was available for users to read through texts in their entirety, the focus of this prototype was on distant reading, allowing a macroanalysis of the corpus through targets queries defined by users. The inclusion of the location view was a later addition to the application, as feedback from colleagues in the Geospatial Humanities team questioned whether the main view alone was sufficient in meeting the requirements of Deep Mapping, given the lack of depth into locations, a sense of connection or meaning was not necessarily created. The location view was intended to address this issue by providing users all of the literary material associated with a location alongside engaging visual media. With the addition of the location view this prototype better met the definition of Deep Mapping, rather than just being a search tool. While this case study focused on distant reading, sectioning out the text based on user queries, the third case study was targeted at close reading, with a focus on allowing users to read one complete text at a time. As described in chapter 2.4.2, this Deep Map fits within the 'prime' category of Ridge et al.'s conceptual model as the concentration on analysis of texts at scale is farther from the concept of spatial narratives in comparison to the careful engagement with a single text, like in close reading practices.

The third case study prototype was the simplest in design out of the four developed for this research. The Reading in Place or Close Reading Deep Map, has a single page view. This view is split into three sections, the top half an interactive panorama, the bottom left a view of the selected text, and the bottom right a map. Users could select any text from the CLDW and it would be displayed in the notebook style text view in the bottom left. As users read through the text they could cycle through the locations mentioned and the map and panoramic views would be updated with the next location. This functionality allows users to see each location they read about in the text, explore the panoramic imagery, and understand the spatial context of the location via the map. As described in chapter 2.4.2, this Deep Map fits within the 'personal' category of Ridge et al.'s conceptual model and given the concentration on close reading of an individual text, is closely related to the idea of a spatial narrative. This prototype provides an engaging interactive experience for the reading of texts that was targeted at the general public (UG3), who might want to enjoy reading a text from the corpus, but also has value to literary scholars performing close reading practices in their research, as this application help put the writings in a spatial context.

The first three case studies all had a focus on the writing from the CLDW and on using web map interfaces as key elements within the prototype. The fourth case study was aimed at moving beyond those constraints to explore a more abstract implementation of Deep Mapping. The site of Tarn Hows was of interest to researchers in the Geospatial Innovations team as it had a significant history of changing landscape and connections to well-known literary authors, specifically Beatrix Potter. In order to explore the creation of a Deep Map could show the history of landscape change at the site and also deviate from the standard 2D web map interface, a 3D solution was attempted. Three scripts were developed for the Unity 3D gaming engine in order to take existing historic and modern 2D maps and automatically create a 3D landscape.

Automation was a key focus of this case study because the manual creation of 3D environments is time consuming and can require significant expertise. Using a manual process would have made this prototype difficult to develop and difficult for other researchers in the future to repeat. The construction of each prototype developed for this thesis was designed to be a usable frame work for other research to build off of in the future, so an automated solution was required for this case study. To create the landscape from an automatic process, three scripts were developed, each focusing on taking data from the 2D sources and transforming them into 3D information. The first script took a 2D elevation map, a grayscale representation of elevation information taken from LiDAR and satellite data, to create the ground geometry of the environment. The result of this script was a colorless terrain that had the geometry of the site. The second script took a 2D vegetation map, derived from Ordinance Survey maps, and for each color in the vegetation map, colored the appropriate part of the terrain with a texture assigned to that type of vegetation, like sand or gravel where water would be, dirt/grass where grass, bushes and trees would be. The last script filled the terrain with 3D vegetation assets, trees, bushes, rocks etc., based on the 2D vegetation map. The combination of these scripts resulted in a complete 3D representation of the landscape as depicted in the two 2D maps used as inputs, as can be seen in Figure 19 and Figure 22. This process was repeated using vegetation and elevation maps that represented different time periods (1850,1900,2018), so that each phase of landscape change was represented in 3D. This case study's prototype used Unity to create an interactive experience where users could explore and compare the 3D landscape in the three time periods. This experience served as a platform for learning and the dissemination of data to all audiences but is somewhat more directed at the general public (UG3). This platform also has significant potential for conservation efforts directed by site authorities (UG2). The ability to compare 3D landscapes from different time periods allows authorities the potential to compare past environments to future planned environments, based on 2D maps of planned changes. Visualizing how planned changes could impact things like iconic vistas can help in conservation efforts and lessen the impact changes make to the public's perception of a site. This can also prove to be a useful tool for historians and archaeologists that are attempting to visualize past environments. 2D vegetation and elevation information are common deliverables of archaeological projects or historic research and could be easily converted to 3D visualizations using the scripts developed for this case study. This Deep Map fits in the middle of the conceptual model proposed by Ridge et al as described in chapter 2.4.2. This Deep Map provides a significant amount of visual information to users, but the scope of the map is not representative of the 'universe' of data. This map provides embedded multimedia to deliver data to users, but the emphasis on exploration in this application moves it away from the side of spatial narratives in the conceptual model. This Deep Map ends up on the edge between 'personal' and 'prime' Deep Maps. The Tarn Hows application performed the best during evaluation, fulfilling all criteria in the evaluation rubric, actively engaging audiences during demonstrations, and providing potential as a viable tool for research and conservation.

The four Deep Map prototypes concentrated solely on the Lake District region, most because that was the subject area of the Geospatial Innovations project and that is where the primary data source was focused. The last case study undertaken for this research was aimed at determining if the processes used to prepare data for the previous case studies was repeatable for external data sources. Given the historic nature and focus on genres like travel journals and poetry, the CLDW is not representative of the wider universe of literary data, so testing the methods used to process the corpus on other sources helps to show that the data, methods, and design of Deep Map prototypes are not exclusive to the Lake District region or CLDW.

The fifth case study was done in cooperation with researchers investigating testimonies of Holocaust survivors. This case study took Holocaust testimonies that had been done two different groups, the United States Holocaust Memorial Museum (USHMM) and the University of Southern California Shoah Foundation Visual History Archive (VHA). The testimonies had been transcribed and digitized but were in unstructured formats and difficult for humans or computers to make sense of for analysis. The first part of this case study was focused on processing the testimonies in order to clean them up, removing elements that were byproducts of the transcription process and had no value for analysis. After cleaning up the transcripts they were passed into an NLP processing tool developed for this thesis as described in chapter 2.2.2. The NLP process was used to identify key elements within the texts that were of interest to the researchers, including personal details, like names and genders, as well as all locations. The locations were then run through the geoparser built for this research and were linked with geographic coordinates via gazetteers from Geonames and ones provided by the researchers that had a specific focus on World War 2 era European locations. All the identified elements were then labeled with XML tags in order to make the information usable for future analysis with corpus linguistic tools or integration into Deep Mapping applications. In this final form the testimonies had been prepared in a manner that matched the processing of the CLDW and would allow for these texts to be directly used in Deep Mapping prototypes from the first four case studies with minor changes.

The final phase of research for this thesis involved the evaluation of the Deep Map prototypes in order to determine their effectiveness. From the previous literature, it was clear that no existing standard evaluation approach existed for Deep Maps, hence an evaluation methodology needed to be created. To provide a broader set of feedback on the prototypes, a survey was created and sent out to the public and academic audiences (representing the three targeted user groups). The survey allowed users to explore the prototypes freely, based on their own interests, and would prompt users with questions as they engaged with different aspects of the maps. The survey, along with feedback from in person demonstration, seminars and conference presentations, was used as one of three categories of evaluation (Data Evaluation). The other two categories used to evaluate the maps were based on the problems and objectives stated for each case study (Design Evaluation) and the research questions of this thesis (Research Evaluation). Each prototype Deep Map was found to fulfil the requirements of being a Deep Map, and each contributed to either the dissemination and engagement of data with the three user groups, or as a tool for informing the academic research process.

8.2 Research Questions

This section looks at how the research questions identified by this thesis, in section 1.5, were addressed by the work undertaken in by this research and the case study prototypes.

> **RQ1** - How can Deep Mapping, in combination with spatial analysis, natural language processing, and corpus linguistic techniques, be used to inform the research process?

This research question is primarily addressed by the Analytic Deep Map, chapter 4, and the Tarn Hows Deep Map, chapter 6. A tool like the Analytic Deep Map, allows for the prototyping of research questions that using traditional methods would be time intensive. Being able to perform a quick comparison between search terms and visualize their spatial distribution can inform researchers of new relationships without requiring a large investment in time. Studies like "Locating the beautiful, picturesque, sublime and majestic (Donaldson 2017)," show that the spatial relationship between words can be significant in the understanding of literature. That type of study would normally require multiple tools to bring the relevant data into a GIS environment in order to see the spatial relationships, all before actually knowing if there is any relationship to see. Using a Deep Map to prototype this type of research saves time and allows for more experimentation in the creation of new research avenues.

The Tarn Hows Deep Map is a viable tool for assisting in visualization of historic environments, which may be of significant interest to literary historians attempting to understand what views might have looked like in the past for locations described in literature. This type of visualization can provide supporting evidence for research activities or inspire new avenues of research. This tool also had significant application for conservation efforts, allowing the planning changes to a site without directly impacting the real-world environment.

The other prototypes developed, like the Reading in Place Deep Map, can be used to help inform the research process, but they provide more of a supportive role, like enhancing the close reading experience, rather than the active exploratory role that the Analytic and Tarn Hows Deep Maps provide.

> **RQ2** - What elements are required to best present historic data to various user groups in a way that makes it accessible and understandable? What visualization techniques best engage these different user groups?

All of the case study prototype contributed to addressing this research question. Through the evaluation survey and in person demonstrations some elements of the maps presented themselves as viable techniques for enabling user accessibility and understanding, and engagement. One of the main elements of accessibility was the use of web applications as the platform each prototype was built on. This provided access to all interested audiences, allowing individuals from different countries to access the applications simultaneously. Web applications are also more accessible and understandable in general for all audiences, most people have some experience using a web browser, while not all users are familiar with the processes of accessing or stating desktop based applications. This difference could be seen in some in person demonstrations of the Tarn Hows map, which was occasionally demonstrated using a desktop version of the platform, which provided better performance on the local machine and was independent of internet connectivity. Users attempting to use the Desktop version had more difficulty in access without intervention from the presenter. With the web applications the only barrier to entry for users was a working browser and the appropriate URL to access the applications.

The understandability of the applications was generally unanimous. Most users have had some exposure to web maps given the prevalence of platforms like Google Maps, Bing Maps and Apple Maps, so the map and marker design of the first three case studies was not difficult for users to engage with. The Tarn Hows Deep Map proved to have the greatest contrast in user understanding and engagement. While all of the users enjoyed the visuals and were engaged by the content, some older users found it difficult to navigate the environment given its dependence on standard 3D video game controls (WASD or arrow keys). Younger audiences actively engaged with this prototype and enjoyed the free exploration and immersion provided by the 3D environment.

The text centric elements, whether in full text form or excerpts were less engaging especially to some of the younger users during demonstrations. The text elements tended to be more engaging for users who had a previous interest in the subject matter. Some users during demonstrations would remember reading a specific text and seek it out in the Deep Map to explore the related material. Engagement was highest with the more visual elements of the prototype applications. The 3D environment, the Panoramic imagery, and the artistic imagery (Drawings and Paintings) tended to engage users from all audiences the most. The visual elements did the best job of drawing in users and keeping them engaged with the applications.

> **RQ3** – How can Deep Mapping applications be evaluated given the vast array of forms a map might take under the loose definitions of Deep Mapping?

This research question was addressed in chapter 8, discussing the evaluation methodology of Deep Mapping using the case study prototypes. Each case study in this research was evaluated using three categories, each with their own set of criteria. While this allowed for each prototype to be evaluated in the context of this thesis, it did not provide significant insight into how all Deep Maps can be evaluated since the definition of Deep Mapping is flexible and the purpose of any Deep Map can vary significantly. The criteria within each category is aimed at each case study in isolation, and while common enough for comparison within this thesis, that criteria would not work for all instances of Deep Maps. The exception to this is in the Data Evaluation category. Most of the criteria in the Evaluation Rubric can be used for the evaluation of Deep Mapping applications outside the context of this thesis. What the evaluation of these case studies did show, is that all Deep Maps should be capable of being evaluated on a higher level, which is the categories themselves. All Deep Maps should be capable of being evaluated based on:

> *Design* - did the application meet the intended design requirements?
>
> *Research* - did the application further the research goals of the project or provide a tool or method for informing future research?
>
> *Data* - does the application fall within the definition of a Deep Map?

By looking at these three aspects of a Deep Map, a high level of evaluation can be used to compare Deep Maps across the field, and the evaluation methodology created in this thesis provides a first version which is flexible enough to be extended for the evaluation of future Deep Maps.

8.3 Contributions

The following contributions can be attributed to the research presented as part of this thesis.

Contribution to the wider field

Deep Mapping is a tool and framework that can be used by researchers to enhance the exploration of research questions using spatial data. Using maps as a means to access texts and other historic data, enables researchers to look at data from a new perspective and potentially present data in more meaningful ways. Deep Mapping provides utility at both ends of the research process. This PhDs research shows various approaches of constructing Deep Maps for research as well as their applications outside the academic context, like conservation applications. This research also provides a practical approach to addressing research questions, using research by design methodologies, and building viable prototypes. The practical nature of this research allows it to bridge the gap between theoretical concepts in the humanities and functional research in the sciences.

Cases to inform intellectual discussion

The work done as part of this PhD and the Geospatial Innovations project has helped inform discourse in the discussions of relationships between literature and maps. The Chronotopic Cartographies[47] is an AHRC-funded research project that emerged from the Geospatial Innovations team, that is investigating how digital tools might be used to analyze, map, and visualize the spaces of literary texts, focusing on real world locations in fictional texts. Discussions of literature and historic landscapes has helped to contribute to a new AHRC-funded network on *Revealing Long-Term Change in Vegetation Landscapes*[48], mapping vegetation described in historic literary sources, using techniques similar to the Analytic Deep Map.

Additionally, the papers published as part of this thesis have been used as citations for others exploring NLP and Deep Mapping issues.

Tool for conservation

The Tarn Hows Deep Map offers site authorities an application that can help visualize past environments to help inform site maintenance, and visualize the future landscape based on planned changes. With sites that have had a rich history of landscape change it can be difficult for authorities to plan maintenance of the site. At Tarn Hows, the modern environment is an iconic tourist destination, but today's landscape differs

[47] https://www.lancaster.ac.uk/chronotopic-cartographies/

[48] https://gtr.ukri.org/projects?ref=AH%2FT006153%2F1

significantly from the past. Authorities must determine what trees can be cut back, to better reflect the historic plantations of the Marshall family, while not removing too much that might negatively affect the picturesque views and visitors' engagement with the natural landscape. The Deep Map provides a means to visualize future plans for the site, allowing authorities to gain feedback on proposed changes without having to modify the real world landscape first.

The National Trust showed interest in this tool for conservation efforts at Tarn Hows and other heritage sites with changing landscapes. Interest in this tool was also shown from archaeologists that attended a university seminar, as 2D vegetation maps are a common output of archaeological projects but they often don't have an effective means of visualizing the collected data. This tool would enable those 2D maps to be used to create 3D visuals. The code for the automatic generation scripts was shared with these archaeologists for their potential future use.

8.4 Future Work

The following would be interesting research avenues to explore in order to extend and improve upon the research presented in this thesis, and to address its limitations.

Unified Framework

Implementations of Deep Maps are still in an exploratory and experimental phase, with little to no common standards apart from the loose definitions associated with Deep Mapping. One potential direction for future work could be the development of a unified framework for building Deep Maps. In the cultural resource management field the Arches Project[49] is an attempt at creating a unified platform for the collection and management of cultural heritage data. Arches aims to provide features that are not region or language dependent and allow for implementation in most situations. The field of cultural resource management (CRM) has the advantage of some standardization in data collection, most archaeological reports include the same types of classification and methods of quantification. Unlike CRM, the data that could be integrated into a Deep Map is unrestricted and had little to no standardization. Chapter 7 shows that different data sources, even though wholly text based like the CLDW, required a layer of processing in order to bring them into a standard usable format. As technology progresses and more data is digitized and annotated it will become easier work with. With more standardized and accessible data a unified framework for Deep Mapping could be a possible path for future research that would

[49] https://www.archesproject.org/

work as a tool to inform research and as a platform for sharing data regardless of region, language and other limiting factors of modern implementations.

Lake District

With the Lake Districts inclusion into the UNECSO World Heritage List, its status as a National Park, and the many heritage based entities, like the National Trust, that manage the region, there are a significant amount of resources that might assist in the creation of an all-encompassing Deep Map of the Lake District. The research undertaken for this PhD focused on specific aspects of the Lake District, and the creation of an all-encompassing Deep Map of the Lake District would have been outside the scope this research could possibly have achieved, but, especially as technology improves, creation of such a Deep Map is not beyond the realm of possibly. Given the various organizations supporting the region and the constant presence of tourism there would seem to be sufficient money and resources available to construct that type of application. An all-encompassing Deep Map of the Lake District could prove an effective tool for educating tourists and the wide public about all aspects of the Lake District and provide potential for researchers to engage with data beyond the scope of their normal research.

Limitations of work.

As with most fields of study, research is limited to the technology of the time. The modern convenience of the web proved a useful platform for the development of the prototype Deep Maps for each case study, and the ease of access to the web allowed for the straightforward dissemination of those applications to all user groups. While the web is a powerful resource, it has significant limits in the amount of data it can handle, based on memory limits, and the types of visualizations and interactions it supports. In the future, with greater capabilities in the web and easier access to linked data, Deep Maps have the potential to develop far beyond what is capable now.

References

Alex B, Byrne K, Grover C, Tobin R. (2015). Adapting the Edinburgh Geoparser for Historical Georeferencing. International Journal of Humanities and Arts Computing. 9. 15-35. 10.3366/ijhac.2015.0136.

Alves, D., Queiroz, A. (2015) Exploring Literary Landscapes: From Texts to Spatiotemporal Analysis through Collaborative Work and GIS. *International Journal of Humanities and Arts Computing*, 9.1: 57-73. Edinburgh University Press.

Ayers, E. L. (2010). Turning toward Place, Space, and Time. In The Spatial Humanities (p. 1). Indiana University Press.

Bellahsene, Zohra & Bonifati, Angela & Duchateau, Fabien & Velegrakis, Yannis. (2011). On Evaluating Schema Matching and Mapping. 10.1007/978-3-642-16518-4_9.

Berman, Lex & Mostern, Ruth & Southall, Humphrey. (2011). On Historical Gazetteers. International Journal of Humanities and Arts Computing. 5. 127-145. 10.3366/ijhac.2011.0028.

Berman, Lex. (2016). Placing Names: Enriching and Integrating Gazetteers..

Biber, D., Finegan, E. and Atkinson, D. 1993. 'ARCHER and its challenges: compiling and exploring a representative corpus of historical English registers', in J. Aarts, P. de Haan and N. Oostdijk (eds.) English Language Corpora: Design, Analysis and Exploitation, pp. 1–13. Amsterdam: Rodopi.

Bodenhamer, D., 2013. The Potential of Spatial Humanities. In: D.J. Bodenhamer, J. Corrigan and T.M. Harris, ed. The Spatial Humanities: GIS and the Future of Humanities Scholarship. Indiana university Press.

Bodenhamer, David & Corrigan, J. & Harris, T.M.. (2015). Deep Maps and Spatial Narratives.

Bodenhamer, David J, Harris, Trevor M, & Corrigan, John. (2013). Deep Mapping and the Spatial Humanities. International Journal of Humanities and Arts Computing, 7(1-2), 170-175.

Carlson J.S. 2016, Romantic Marks and Measures: Wordsworths poetry in fields of print. University of Pennsylvania Press, Philadelphia

Charniak E. Statistical Language Learning. Cambridge, MA: The MIT Press; 1996.

Cooper D. and Gregory I.N. (2011) "Mapping the English Lake District: A literary GIS" *Transactions of the Institute of British Geographers*, 36, pp. 89-108. See DOI: 10.1111/j.1475-5661.2010.00405.x

Cresswell, T., ed., 2015. *Place: An Introduction.* Wiley- Blackwell. West Sussex.

Cronin, N., 2016. The fineness of things: The Deep Mapping Projects of Tim Robinson's Art and Writings, 1969–72. In: D. Gladwin and C. Cusick, ed. 2016. *Unfolding Irish Landscapes: Tim Robinson, culture and environment.* Manchester university Press.

Darrel J. Rohl, 'Chorography: History, Theory and Potential for Archaeological Research/ in Duggan, M., F. McIntosh and D.J. Rohl (eds) TRAC 2011: Proceedings of the Twenty First Theoretical Roman Archaeology Conference (Oxford: Oxbow Books, 2012) 19-32.

Diamantini C., Mircoli A., Potena D. and Storti E., Semantic disambiguation in a social information discovery system, 2015 International Conference on Collaboration Technologies and Systems (CTS), Atlanta, GA, 2015, pp. 326-333, doi: 10.1109/CTS.2015.7210442.

Donaldson C., Gregory I.N. and Taylor J.E. (2017) "Locating the beautiful, picturesque, sublime and majestic: Spatially analysing the application of aesthetic terminology in descriptions of the English Lake District" *Journal of Historical Geography*, 56, pp. 43-60

dos Santos C.N., Milidiú R.L. (2012) Named Entity Recognition. In: Entropy Guided Transformation Learning: Algorithms and Applications. SpringerBriefs in Computer Science. Springer, London. https://doi-org.ezproxy.lancs.ac.uk/10.1007/978-1-4471-2978-3_7

Dutt S, Chandramouli S. 2018. Machine Learning. Pearson India

Eekels, J., and Roozenburg, N.F.M. A methodological comparison of the structures of scientific research and engineering design: Their similarities and differences. Design Studies, 12, 4 (1991), 197–203.

Ethington Philip J., MacDonald Beau, Stein Gary, Deverell William, and Longcore Travis (2020) Historical Ecology of the Los Angeles River Watershed and Environs, Report to the John Randolph Haynes and Dora Haynes Foundation

Findeli, Alain & Brouillet, Denis & Martin, Sophie & Moineau, Christophe & Tarrago, Richard. (2008). Research Through Design and Transdisciplinarity: A Tentative Contribution to the Methodology of Design Research.

Fisher Crosthwaite J.'Peter Crosthwaite: The Founder of Crosthwaite's Museum, Keswick', Transactions of the Cumberland Association for the Advancement of Literature and Science, Part III (1778), 151¬–64.

Fishkin, S. (2011) "Deep Maps": A Brief for Digital Palimpsest Mapping Projects (DPMPs, or "Deep Maps"). *Journal of Transnational American Studies*, 3(2) 1-31.

Fries, C. 1940. American English Grammar. New York: Appleton Century.
1952. The Structure of English. London: Longman.

Godin, D., & Zahedi, M. (2014). Aspects of Research through Design : A Literature Review.

Gonzalo Navarro. 2001. A guided tour to approximate string matching. ACM Comput. Surv. 33, 1 (March 2001), 31–88.
DOI:https://doi.org/10.1145/375360.375365

Gregory I.N. and Cooper D. (2009) "Thomas Gray, Samuel Taylor Coleridge and Geographical Information Systems: A Literary GIS of Two Lake District Tours" *International Journal of Humanities and Arts Computing*, 3, pp. 61-84.
See DOI: 10.3366/ijhac.2009.0009

Gregory I.N**.**, Donaldson C., Murrieta-Flores P. and Rayson P. (2015) "Geoparsing, GIS and textual analysis: Current developments in Spatial Humanities research" *International Journal of Humanities and Arts Computing*, 9, pp. 1-14.
See: DOI: 10.3366/ijhac.2015.0135

Gregory, Ian, Donaldson, Christopher, Murrieta-Flores, Patricia, & Rayson, Paul. (2015). Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research. International Journal of Humanities and Arts Computing, 9(1), 1-14.

Gregory,I.N.,Cooper,D.,Hardie,A.,Rayson,P.,2015.Spatializing and analysing digital texts:corpora,GIS and places. In:Bodenhamer,D.,Corrigan,J.,Harris,T. (Eds.), Spatial Narratives and DeepMaps. Indiana University Press, Bloomington, pp.150–178.

Grover Claire, Tobin Richard, Byrne Kate, Woollard Matthew, Reid James, Dunn Stuart and Ball Julian 2010 Use of the Edinburgh geoparser for georeferencing digitized historical collectionsPhil. Trans. R. Soc. A.3683875–3889 http://doi.org/10.1098/rsta.2010.0149

Harris, T., Corrigan, John, & Bodenhamer, David J. (2010). The Spatial Humanities GIS and the Future of Humanities Scholarship (Spatial humanities). Bloomington: Indiana University Press.

Heat, M. W. L. (1991). *PrairyErth: (a deep map).*

Hevner, A.R.; March, S.T.; and Park, J. Design research in information systems research. MIS Quarterly, 28, 1 (2004), 75–105.

Honn, Josh. "A Guide to Digital Humanities: Values Methods". Northwestern University Library. Archived from the original on 19 September 2015. Retrieved 19 September 2015.

Kennedy-clark, Shannon. Research by Design: Design-Based Research and the Higher Degree Research student. Journal of Learning Design, [S.l.], v. 6, n. 2, p. 26-32, sep. 2013. ISSN 1832-8342. Available at: <https://www.jld.edu.au/article/view/128>. Date accessed: 02 Aug. 2017. doi:http://dx.doi.org/10.5204/jld.v6i2.128.

Knight, D., Evans, D., Carter, R. and Adolphs, S. 2009. 'HeadTalk, HandTalk and the corpus: towards a framework for multi-modal, multi-media corpus development', Corpora 4 (1): 1–32.

Knowles, A.K. and Hillebrand, J. with Jaskot, P.B. and Walke, A. 2020. Integrative, Interdisciplinary Database Design for the Spatial Humanities: the Case of the Holocaust Ghettos Project. International Journal of Humanities and Arts, 14(1-2), pp. 64-80.

Kripke, Saul. 1982. Naming and Necessity. Boston: Harvard University Press.

Kurwakumire, E. (2014) Towards a Public Sector GIS Evaluation Methodology. South African Journal of Geomatics, Vol. 3, No. 1, January 2014

Leech, G. 1971. Meaning and the English Verb. London: Longman.
2007. 'New resources, or just better old ones?', in M. Hundt, N. Nesselhauf and C. Biewer (eds.) Corpus Linguistics and the Web, pp. 134–49. Amsterdam: Rodopi.

Lunenfeld, Peter, Drucker, Johanna, Schnapp, Jeffrey, Burdick, Anne, & Presner, Todd. (2012). Digital_Humanities. Cambridge: The MIT Press Open.

Maher, S. (2005) Deep Mapping History: Wallace Stegner's Wolf Willow and William Least Heat-Moon's PrairyErth: A Deep Map. Heritage of the Great Plains, V.38 No.1, 39-54.

Manning CD, Schutze H. Foundations of Statistical Natural Language Processing. The MIT Press; 2000.

Martinez, A.R. (2010), Natural language processing. WIREs Comp Stat, 2: 352-357. doi:10.1002/wics.76

McDonald, David D. 1993. Internal and External Evidence in the Identification and Semantic Categorization of Proper Names. In Proc. Corpus Processing for Lexical Acquisition.

McDonough, Katherine, Moncla, Ludovic, & Van de Camp, Matje. (2019). Named entity recognition goes to old regime France: Geographic text analysis for early modern French corpora. International Journal of Geographical Information Science, 33(12), 2498-2522.

McEnery, Tony, and Hardie A. *Corpus Linguistics : Method, Theory and Practice*, Cambridge University Press, 2011. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=807166.

Megahed Y (2017) On research by design. *Arg.* 21.4 338-343. Cambridge University Press 2018

Murrieta-Flores P, Gregory I, Liceras-Garrido R, Bellamy K, Martins B. Theorising the Spatial Humanities. Digital Humanities 2019 Utrecht University.

Murrieta-Flores, P.,Baron,A.,Gregory,I.N.,Hardie,A.,Rayson,P.,2015.Automatically analysing large texts in a GIS environment: The Registrar General's reports and cholera in the nineteenth century. Transactions in GIS 19;,pp. 296–320. http://dx.doi.org/10.1111/tgis.12106.

Naramore Maher, S. 2014. *Deep Map Country: Literary Cartography of the Great Plains.* Nebraska.

Oxx, K., Brimicombe, A., Rush, J. (2013) Envisioning Deep Maps: Exploring the Spatial Navigation Metaphor in Deep Mapping. *International Journal of Humanities and Arts Computing*, 7.1-2: 201-227. Edinburgh University Press.

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

Peffers, K., Tunnanen, T., Rothenberger M., Chatterjee S. (2007) A Design Science Research Methodology for Information Systems Research. *Journal of Management Information System*, 24:3 45-77.

Poibeau, Thierry; Kosseim, L. 2001. Proper Name Extraction from Non-Journalistic Texts. In Proc. Computational Linguistics in the Netherlands.

Porter C. Introduction: The Importance of Place and Openness in Spatial Humanities Research. (2018). International Journal of Humanities and Arts Computing, 12(2), 91-101.

Porter, Catherine, Atkinson, Paul, & Gregory, Ian. (2015). Geographical Text Analysis: A new approach to understanding nineteenth-century mortality. Health & Place, 36, 25-34.

Presner, T., Shepard, D., & Kawano, Y. (2014). Hypercities Thick Mapping in the Digital Humanities. Retrieved from https://escholarship.org/uc/item/3mh5t455

Rayson, P. 2008 From key words to key semantic domains. International Journal of Corpus Linguistics. 13, 4, p. 519-549. 31 p.

Rayson, P. E., **Reinhold, A.**, Butler, J., Donaldson, C. E., Gregory, I. N. & Taylor, J. E., 7/11/2017, A deeply annotated testbed for geographical text analysis: The Corpus of Lake District Writing. *GeoHumanities'17 Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities.* New York: Association for Computing Machinery (ACM), p. 9-15 7p.

Reinhold, A., Donaldson, C. E., Gregory, I. N. & Rayson, P. E., 14/09/2017, Exploring Deep Mapping Concepts: Crosthwaite's Map and West's Picturesque Stations. *Proceedings of Workshops and Posters at the 13th International Conference on Spatial Information Theory* (COSIT 2017). Fogliaroni, P., Ballatore, A. & Clementini, E. (eds.). Cham: Springer, p. 265-273 9 p. (Lecture Notes in Geoinformation and Cartography).

Reinhold, A., Gregory, I. N. & Rayson, P. E., 2018, Deep Mapping Tarn Hows: Automated Generation of 3D Historic Landscapes. *Eurographics Workshop on Graphics and Cultural Heritage.* Sablatnig, R. & Wimmer, M. (eds.). The Eurographics Association

Ridge, M., Lafreniere, D., Nesbit, S. (2013) Creating Deep Maps and Spatial Narratives Through Design. *International Journal of Humanities and Arts Computing*, 7.1-2: 176-189. Edinburgh University Press.

Roggema, R. Research by Design: Proposition for a Methodological Approach. Urban Sci. 2017, 1, 2.

Rossi, M., and Sein, M.K. Design research workshop: A proactive research approach. Paper presented at the Twenty-Sixth Information Systems Research Seminar in Scandinavia, Information Systems Research in Scandinavia Association, Haikko, Finland, August 9–12, 2003.

Saar M and Palang H, "The Dimensions of Place Meanings", Living Rev. Landscape Res., 3 (2009), 3, doi:10.12942/lrlr-2009-3, URL (accessed <date>): http://lrlr.landscapeonline.de/lrlr-2009-3.

Schofield T, Whitelaw M, Kirk D, Research through design and digital humanities in practice: What, how and who in an archive research project, Digital Scholarship in the Humanities, Volume 32, Issue suppl_1, April 2017, Pages i103–i120, https://doi.org/10.1093/llc/fqx005

Sekine, S., & Ranchhod, Elisabete. (2009). *Named entities recognition, classification, and use* (Benjamins current topics ; v. 19). Amsterdam ; Philadelphia: John Benjamins Pub. Company.

Silveira, Luis. (2014). Geographic Information Systems and Historical Research: An Appraisal. International Journal of Humanities and Arts Computing. 8. 28-45. 10.3366/ijhac.2014.0118.

Sinclair, J. (1991). Corpus, concordance and collocation. Oxford: Oxford University Press.

Thielen, Christine. 1995. An Approach to Proper Name Tagging for German. In Proc. Conference of European Chapter of the Association for Computational Linguistics. SIGDAT.

Thrift, N. J. (2006). Space, Place, and Time. In The Oxford Handbook of Contextual Political Analysis (Vol. 1, Oxford Handbooks of Political Science, pp. The Oxford Handbook of Contextual Political Analysis, 2006-03-16, Vol.1). Oxford University Press.

Tognini-Bonelli, Elena. (2002). Corpus Linguistics at Work. Computational Linguistics. 28. 583-583. 10.1162/coli.2002.28.4.583a.

Tuan Y. 1979 Space and Place The Perspective of Experience. (1979). University of Minnesota Press.

Walls, J.; Widmeyer, G.; and El Sawy, O. Building an information system design theory for vigilant EIS. Information Systems Research, 3, 1 (1992), 36–59.

West T, A guide to the Lakes: Dedicated to the lovers of landscape studies, and to all who have visited, or intend to visit, the lakes in Cumberland, Westmorland, and Lancashire. Printed for Richardson and Urquhart, under the Royal Exchange, and W. Pennington, Kendal, 1778.

Westerveld, L. and A.K. Knowles. 2020. Loosening the grid: topology as the basis for a more inclusive GIS. International Journal of Geographical Information Science. Advance online publication.

Woodsworth W, Directions and Information for the Tourist Directions and Information for the Tourist. 1835

Zimmerman J, Forlizzi J, and Evenson S. 2007. Research through design as a method for interaction design research in HCI. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’07). Association for Computing Machinery, New York, NY, USA, 493–502. DOI:https://doi.org/10.1145/1240624.1240704